



# Mission-Based Test and Evaluation Assessment Process Guidebook

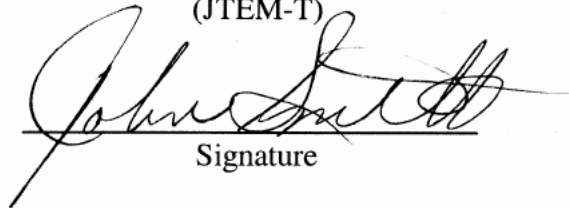


**April 1, 2011**

---

Approved By:  
John Smith  
Director

Joint Test and Evaluation Methodology - Transition  
(JTEM-T)

A handwritten signature in black ink, appearing to read "John Smith", is written over a horizontal line.

Signature

DISTRIBUTION STATEMENT A: Approved for public release: distribution unlimited.

This page intentionally left blank.

## FOREWORD

This *Mission-Based Test and Evaluation Assessment Guidebook* (hereafter referred to as “guidebook”) responds to an operational test agency stated need for a documented analysis framework and serves as a complementary process to the mission decomposition methodology described in the *Measures Development Standard Operating Procedure (SOP)*, Version 2, published in January 2011. As the SOP offers practical advice on how to develop measures of system and system-of-systems (SoS), this guidebook provides a means to evaluate a system’s impact on task performance and mission effectiveness given relevant attributes and measures at the system, task, and mission levels. As a guide, this document provides the following:

- A process to assess system/SoS attributes to include key performance parameters, key system attributes, and other system/SoS attributes
- A methodology to assess system impact on task performance and mission effectiveness
- A means to assess the risk of an incorrect assessment
- A disciplined and repeatable process for assessing mission accomplishment, otherwise known as combat mission effectiveness

This guidebook does not focus on the evaluation of individual measures since practical handbooks and statistical methodologies already exist that do so. Instead, this guidebook offers a methodology to assess system/SoS functional attributes, task performance, and mission effectiveness across numerous measures of effectiveness, performance, and suitability to include an assessment of the level of confidence (risk) in making a conclusion based on the measures and assessment process.

The assessment process included in this guidebook offers flexibility to the evaluator through the use of various scoring and assessment models that are selected based on need. Additionally, the risk model, color coding models, and weighting criteria can all be adjusted to suit the needs of the user.

An electronic version of this guidebook, along with the *Measures Development SOP*, is available on the unclassified Defense Acquisition University Acquisition Community Connection website at <https://acc.dau.mil/TIJE>.

This page intentionally left blank.

## TABLE OF CONTENTS

CHAPTER 1 MISSION-BASED TEST AND EVALUATION .....	1-1
INTRODUCTION .....	1-1
KEY CONCEPTS IN ASSESSMENT .....	1-2
Measures Framework .....	1-2
Assessment Levels .....	1-3
Conditions .....	1-3
Measures .....	1-4
Measure Values .....	1-5
Aggregating Measures .....	1-6
Prioritization and Weighting .....	1-6
Frame of Reference – Scoring Models .....	1-6
Measurement Error .....	1-7
SCOPING THE ASSESSMENT MODEL .....	1-7
Element Descriptions .....	1-7
GUIDEBOOK PLAN OF ACTION .....	1-10
CHAPTER 2 ASSESSMENT PROCESS .....	2-1
STEP 1: INITIATING THE ASSESSMENT PROCESS .....	2-1
Relationship Mapping .....	2-1
Prioritization .....	2-3
STEP 2: CONDUCT THE ASSESSMENT .....	2-12
System/SoS Assessment .....	2-15
Task Performance Assessment .....	2-20
Mission Effectiveness Assessment .....	2-25
STEP 3: REPORTING STANDARD .....	2-27
CHAPTER 3 ASSESSMENT PROCESS EXAMPLE .....	3-1
STEP 1: INITIATING THE ASSESSMENT PROCESS .....	3-1
Relationship Mapping .....	3-1
Prioritization .....	3-3
STEP 2: CONDUCT THE ASSESSMENT .....	3-10
System/SoS Assessment .....	3-10
Task Performance Assessment .....	3-14
Mission Effectiveness Assessment .....	3-18
STEP 3: REPORTING STANDARD .....	3-20
CHAPTER 4 SUMMARY .....	4-1
INTRODUCTION .....	4-1
FOCUS .....	4-1
FLEXIBILITY .....	4-2
ENABLING DESIGN OF EXPERIMENTS .....	4-2

## LIST OF FIGURES

Figure 1-1. Measures Development and Assessment Process .....	1-2
Figure 1-2. Measure Observed vs. Planned Scale .....	1-5
Figure 1-3. Simple Task Model .....	1-8

Figure 1-4. Complex Task Model (Example 1) .....	1-9
Figure 1-5. Complex Task Model (Example 2) .....	1-9
Figure 2-1. Prioritization Design Space .....	2-3
Figure 2-2. Prioritizing and Weighting Components .....	2-7
Figure 2-3. Risk Matrix .....	2-14
Figure 2-4. System – Task Causality Diagram .....	2-31
Figure 3-1. System – Task Causality Diagram .....	3-25

## LIST OF TABLES

Table 2-1. Required Relationships for Mission-Based Assessment .....	2-2
Table 2-2. Example Relationship Matrix (Task Attribute vs. Mission Attribute) .....	2-2
Table 2-3. Mission – Condition Weighting Example Format .....	2-8
Table 2-4. Mission Effects Weighting Example Format .....	2-9
Table 2-5. Task Performance Weighting Example Format .....	2-10
Table 2-6. Task Weighting Schema .....	2-11
Table 2-7. System/SoS Attribute Weighting Example Format .....	2-11
Table 2-8. System/SoS Attribute Weighting Schema .....	2-12
Table 2-9. Measure Color Codes .....	2-13
Table 2-10. Risk Color Codes .....	2-14
Table 2-11. Levels of Likelihood - Reliability and Validity .....	2-15
Table 2-12. Impact Levels .....	2-15
Table 2-13. System/SoS Scoring Table .....	2-17
Table 2-14. System/SoS Risk Scoring Table .....	2-18
Table 2-15. Aggregate System/SoS Scoring Table .....	2-19
Table 2-16. Aggregate System/SoS Risk Scoring Table .....	2-19
Table 2-17. System/SoS Attribute Exceptions .....	2-20
Table 2-18. Task Performance Scoring Table .....	2-21
Table 2-19. Task Performance Risk Scoring Table .....	2-23
Table 2-20. Aggregate Task Performance Scoring Table .....	2-24
Table 2-21. Aggregate Task Performance Risk Scoring Table .....	2-24
Table 2-22. Task Performance Exceptions .....	2-25
Table 2-23. Sample Format for a Qualitative Assessment of Mission Effects .....	2-26
Table 2-24. Sample System/SoS Assessment Report .....	2-28
Table 2-25. Sample Task Performance Assessment Report .....	2-29
Table 2-26. Mission - Task - System Linkages .....	2-30
Table 2-27. Root Cause Analysis .....	2-32
Table 3-1. Required Relationships for Mission-Based Assessment .....	3-1
Table 3-2. Condition Sets for JPR Mission .....	3-2
Table 3-3. Task Attributes to Mission Attributes Relation .....	3-2
Table 3-4. Required Relationships Prioritizations .....	3-3
Table 3-5. Condition Sets Weightings for JPR Mission .....	3-3
Table 3-6. JPR Desired Effects Weight Calculations .....	3-4
Table 3-7. JPR Mission Attribute Weights .....	3-4
Table 3-8. JPR Mission Measure Weights .....	3-5
Table 3-9. “Locate” Tasks Weights .....	3-6

Table 3-10. Task Attribute Weights.....	3-6
Table 3-11. JPR “Locate” Task Measure Weights .....	3-7
Table 3-12. Condition Sets Weightings for SUT Attributes.....	3-8
Table 3-13. System/SoS Attribute Weighting Schema.....	3-8
Table 3-14. System/SoS Measures Weighting Schema.....	3-9
Table 3-15. Assessment Models .....	3-10
Table 3-16. PLB System/SoS Scoring Table.....	3-11
Table 3-17. PLB System/SoS Risk Scoring Table.....	3-12
Table 3-18. Aggregate System/SoS Scoring Table.....	3-13
Table 3-19. Aggregate System/SoS Risk Scoring Table .....	3-13
Table 3-20. System/SoS Attribute Exceptions.....	3-14
Table 3-21. Task Performance Scoring Table .....	3-15
Table 3-22. Task Performance Risk Scoring Table .....	3-16
Table 3-23. Aggregate Task Performance Scoring Table.....	3-17
Table 3-24. Aggregate Task Performance Risk Scoring Table .....	3-17
Table 3-25. Task Performance Exceptions Report .....	3-18
Table 3-26. Mission COI and Sub-COIs.....	3-18
Table 3-27. Tasks to Mission Desired Effects Relationships .....	3-19
Table 3-28. Mission Assessment .....	3-20
Table 3-29. PLB System/SoS Assessment Report.....	3-21
Table 3-30. PLB Impacts on Task Performance Assessment Report .....	3-22
Table 3-31. Mission Assessment .....	3-23
Table 3-32. Mission – Task – System Linkages .....	3-24
Table 3-33. Root Cause Analysis.....	3-26

## LIST OF ANNEXES

Annex A. Acronyms and Abbreviations .....	A-1
Annex B. Definitions of Terms.....	B-1
Annex C. Measures Development Example Matrices .....	C-1
Annex D. Assessment Process Scoring Models .....	D-1
Annex E. Assessment Process Mathematical Models .....	E-1

This page intentionally left blank.



## CHAPTER 1

### MISSION-BASED TEST AND EVALUATION

#### INTRODUCTION

This *Mission-Based Test and Evaluation Assessment Guidebook* (hereafter referred to as “guidebook”) provides a practical methodology and example for assessing a system under test (SUT) and its impact on task performance and mission effectiveness. This is based on the measures framework found in the Joint Test and Evaluation Methodology (JTEM) Capability Test Methodology and the mission decomposition process documented in the *Measures Development Standard Operating Procedure* (SOP), Version 2, published in January 2011. Familiarity with these documents before using this guidebook will enable the user to fully understand the relationship of system and system-of-systems (SoS) attributes and measures, the impact on task performance, and the assessment of mission effectiveness. This guidebook uses the joint personnel recovery (JPR) mission thread and the personal locator beacon (PLB) system example found in the SOP. Refer to annex C of this guidebook for example measures development matrices.

Mission-Based Test and Evaluation (MBT&E) is a concept that was born with the advent of the Joint Capabilities Integration and Development System (JCIDS). Although not formally defined by the Department of Defense (DoD), MBT&E is based on the JCIDS definition of “operational effectiveness.”

#### **Definition of “Operational Effectiveness”**

Operational effectiveness is a measure of the overall ability of a system to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, tactics, supportability, survivability, vulnerability, and threat.

(Chairman, Joint Chiefs of Staff Instruction [CJCSI] 3170.01G, Joint Capabilities Integration and Development System [JCIDS], March 1, 2009)

The Defense Acquisition Process builds on the idea of measuring mission accomplishment by specifying that, “T&E [test and evaluation] should be used to assess improvements to mission capability and operational support based on user needs and should be reported in terms of operational significance to the user.”<sup>1</sup> The Army Test and Evaluation Command (ATEC) defines MBT&E as “a methodology that focuses T&E on the capabilities provided to the warfighter.”<sup>2</sup> Measuring mission and task performance enables the analyst to answer the warfighters’ questions by describing how individual system performance affects the end-state performance of the SoS.

<sup>1</sup> Department of Defense Instruction (DoDI) 5000.02p, *Operation of the Defense Acquisition System*, Enclosure 2,

<sup>2</sup> United States Army Test and Evaluation Command (ATEC) Interim Policy Guidance 10-5, *Mission-Based Test and Evaluation (MBT&E)*, Dated May 17, 2010

Mission and task measures are developed to evaluate military capabilities impact on warfighter effects. The Measures Development SOP describes a process for developing those measures based on terms and concepts found in JCIDS, the DoD acquisition process, joint publications, DoD Architecture Framework (DoDAF) products, and other authoritative sources. The SOP provides a complete end-to-end process of decomposing mission and tasks into attributes and measures, and then tracing system attributes and measures to task performance and mission effectiveness. Figure 1-1 illustrates this process as a systems engineering “V” diagram and will be referred to as the “T&E-V.” The left side of the T&E-V represents the decomposition process, and the right side of the T&E-V represents the assessment process. At the base of the T&E-V is the test design process that determines how the test will be conducted to gather the data necessary to do the evaluation. The SOP describes the process for the left side of the T&E-V. This guidebook addresses the right side of the T&E-V.

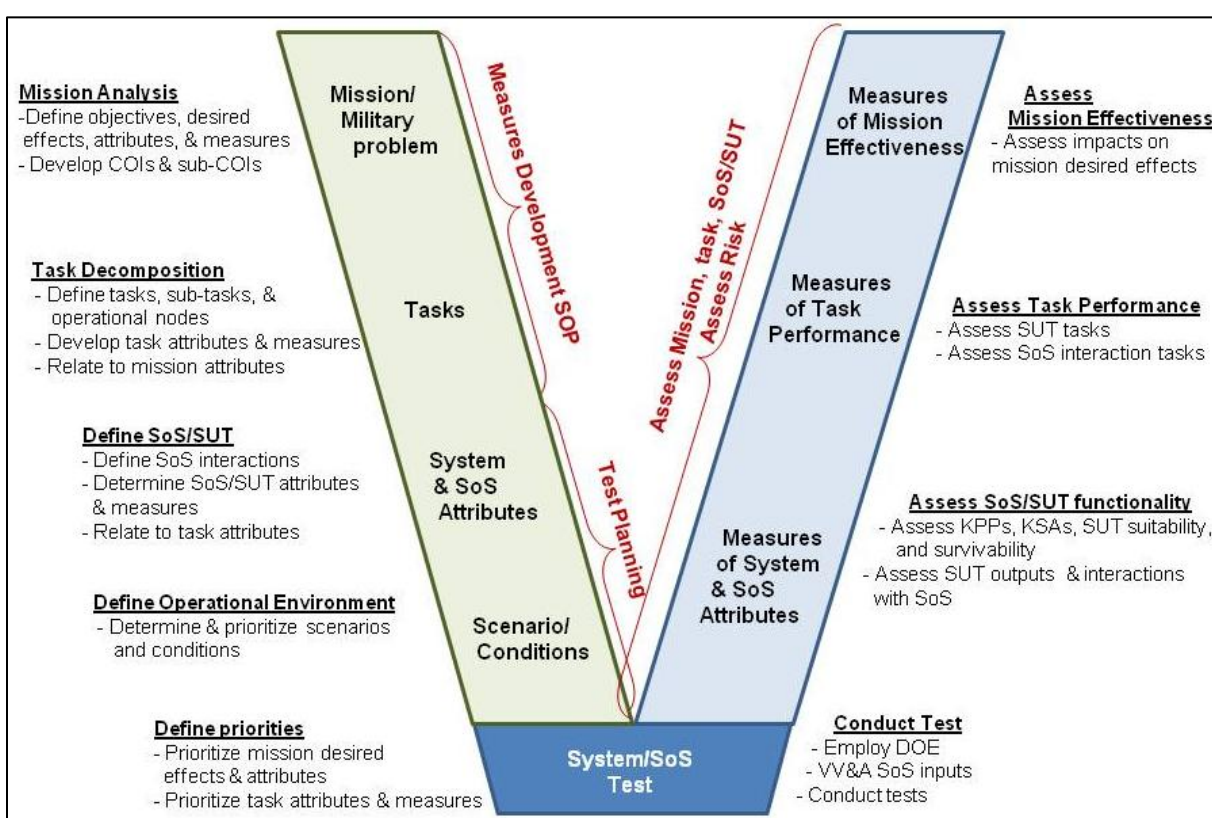


Figure 1-1. Measures Development and Assessment Process

## KEY CONCEPTS IN ASSESSMENT

### Measures Framework

The measures framework is anchored to the JCIDS definition of “capability.” The framework is a relationship diagram of capability key elements (that is, means and ways, desired effects, tasks, standards, and conditions) that identifies the basic questions of who, what, why, and how and then connects measures to each. The measures framework relies on a lexicon derived from joint sources, provides a logical framework for identifying measures in a joint mission environment,

and enables a traceability of measures back to capability requirements. For a more in depth understanding of the measures framework and how it is constructed, refer to the *Measures Development SOP*.

### **Assessment Levels**

The measures framework is characterized as having the following three levels of assessment: mission level to assess mission effectiveness, task level to assess task performance, and system/SoS level to assess system/SoS functions. Mission and task measures are focused on evaluating “how well” a capability performs tasks and achieves mission desired effects. System/SoS measures are focused on “how capable” the system/SoS are in terms of functionality and technical design. Mission measures are generally associated with a SoS. However, decomposing the mission into tasks and then selecting relevant segments of the mission thread will enable a focus at a system level that supports the overall SoS. System measures tend to focus on system-specific attributes that enable assessment of system functionality. System attributes are typically described as key performance parameters (KPP), key system attributes (KSA), and other system attributes.

Based on the previous discussion, one might ask, “Does meeting task performance criteria result in mission effectiveness?” Not necessarily!

What may result is a codependence among different characteristics of the system or a “confounding effect” caused by a spurious relationship across system variables. Confounding is a threat to the validity of inferences made about cause and effect in that a confounding variable (third variable) can adversely affect the relationship between the independent variable and the dependent variable. This could lead to an incorrect conclusion of cause and effect. Thus, to assume that if a system functions according to specifications, then it will perform tasks successfully, is not a valid assumption. Equally, assuming a system or SoS performs tasks successfully does not imply mission effectiveness.

### **Conditions**

The Defense Acquisition Process specifies that “OT&E [Operational Test and Evaluation] shall be used to determine the operational effectiveness and suitability of a system under realistic operational conditions ... .”<sup>3</sup> A “condition” is defined as those variables of an operational environment or situation in which a unit, system, or individual is expected to operate and may affect performance (Joint Publication 1-02, *Department of Defense Dictionary of Military and Associated Terms*). Thus, conditions are independent variables that may impact SUT performance. Descriptors are values of the condition. For example, if ambient temperature is a condition, two descriptors may be below freezing and above freezing. The *Universal Joint Task List (UJTL) Manual* categorizes conditions as descriptors of the physical environment, military environment, or civil environment.<sup>4</sup>

---

<sup>3</sup> DODI 5000.02p, *Operation of the Defense Acquisition System*, Enclosure 6, Page 5, Dated December 8, 2008

<sup>4</sup> Chairman, Joint Chiefs of Staff Manual (CJCSM) 3500.04E, *Universal Joint Task Manual*, Enclosure C, Dated August 25, 2008

### **Example**

Given a maritime environmental condition of surface or sub-surface use, a water resistant watch may be found operationally effective in high moisture areas for those standing watch on a ship. However, if the watch is intended to be used by divers under water, then the watch will probably not be operationally effective, and a waterproof watch may be necessary.

The *Universal Joint Task Manual*<sup>5</sup> describes conditions further as:

- Conditions should be factors of the immediate environment.
- Conditions should directly affect the performance of a task. A condition must directly affect the ease or difficulty of performing at least one task.
- Conditions should not be a related task.
- Each condition should have a unique, understandable name.
- Conditions may apply to all Levels of War and all types of tasks.
- Conditions and descriptors should be written to be compatible with task, conditions, standards framework. Conditions are expressed within the framework of the phrase, “perform this task under conditions of...”. Therefore, each condition and condition descriptor phrase should fit within this framework.

To consider each condition and its descriptors as a separate test factor would not be feasible. Consider a simple problem of three conditions with two descriptors each.<sup>6</sup> The set of combinations will be  $(2^3) = 8$ . Just including an additional three conditions with two descriptors each would then result in a full factorial design requirement of 64 combinations. This can easily overwhelm test resources and require sophisticated design of experiment techniques to minimize test requirements. This guidebook does not discuss design of experiments (DOE) options; however, the Capability Test Methodology Analyst Handbook, Annex D, is a good resource for DOE tools and techniques. The important point is that conditions should be thought of in terms of sets of conditions that help describe a scenario in which the system or SoS will perform. Each scenario should be prioritized based on the probability of the SUT being used within that set of conditions. Those scenarios and sets of conditions with lowest probability of occurrence may not be included in the test design based on resource constraints.

### **Measures**

Measures are developed around attributes. Attributes, as defined by JCIDS, are characteristics of things or activities that can be measured for quality or quantity. Since mission and tasks are defined as activities, and systems are defined as things, attributes apply to each. The measures development process in the Measures Development SOP identifies attributes at the mission, task, and system levels by examining mission, task, and system descriptions through JCIDS documents. Additional sources (joint doctrine, joint publications, future joint concepts, an analytic agenda, and so forth) can be used, as available, if more detail is needed for the measures development process.

---

<sup>5</sup> CJCSM 3500.04E, *Universal Joint Task Manual*, Dated August 25, 2008

<sup>6</sup> Ibid (Calls for three descriptors for each test factor)

A measure is used to estimate the value of an attribute. Estimation involves a certain level of error and must be considered in the evaluation of a measure.

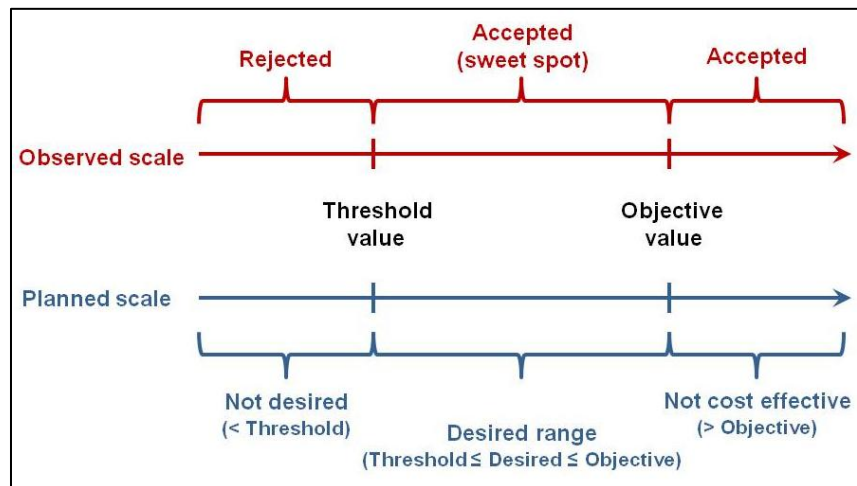
Error will be based on three properties of a measure: reliability, validity, and sensitivity.<sup>7</sup>

- Reliability is the extent to which the measure produces the same result when used repeatedly to measure the same thing under the same conditions
- Validity is the extent to which the measure succeeds at measuring what it is intended to measure
- Sensitivity is the extent to which the values of the measure change when a change or difference occurs in the thing being measured

Reliability and validity should be considered in constructing the measure and means for collecting data. Sensitivity will be based on the value of the independent variable and rate of change. Statistical analysis is a good way to address sensitivity issues.

### Measure Values

A measure value provides the means to determine the independent-dependent variable relationship. Whether the measure is quantitative or qualitative, data is collected and analyzed for a measure to produce an “observed” value that can then be compared to its planned value based on desired threshold and objective values. (See figure 1-2.) A threshold value is defined as a minimum acceptable operational value below which the utility of the system becomes questionable (JCIDS Manual). An objective value is defined as the desired operational goal associated with a performance attribute beyond which any gain in utility does not warrant additional expenditure (JCIDS Manual). The objective value is usually an operationally significant increment above the threshold, but at times may be the same as the threshold when an increment above the threshold is not significant or useful.



**Figure 1-2. Measure Observed vs. Planned Scale**

<sup>7</sup> Rossi, Peter H., Lipsey, Mark W., and Freeman, Howard E., *Evaluation: A Systematic Approach*, Seventh Edition, Sage Publications, Inc.: Thousand Oaks, CA, 2004

A measure consists of a measure “description” and a “scale.”<sup>8</sup> The description defines the attribute that it will measure. The scale establishes a level of reference to compare values. A scale may be a linear numerical scale, or it may be based on an ordinal scale (rejected, accepted, and so forth).

### **Aggregating Measures**

Aggregating measures into a single value presents certain challenges that must be considered. This involves understanding the relative importance of each measure and developing a common frame of reference.

### **Prioritization and Weighting**

The relative importance of a measure may be determined by prioritization and weighting of mission, task, attributes, and measures. For example, if the SUT is expected to operate in one scenario and set of conditions 75% of the time, then logically the assessment of that SUT in a realistic operational environment should be weighted as 75% mission effective if the SUT supports meeting those mission desired effects.

Prioritization is the act of listing or rating items in order of priority. Weighting is a means to determine the relative importance of each item in a list of like things. Weighting is an assignment of a numeric weight where the higher the weight, the higher the importance. Weighting may be based on a probability of occurrence, desired outcome, maximum payoff value, level of accepted risk to the decision-maker, and so forth. It is expected that a higher priority item will have a higher weight assigned to it.

Prioritization and weighting are common practices in decision theory.<sup>9</sup> They are used to calculate expected value or utility in decision analysis. Several statistical tools and methods are available to organize evidence, evaluate risks, and aid in decision-making. However, due to uncertainties in outcomes, lack of quantifiable data, or levels of risk aversion, priorities and weightings may need to be assessed through judgment and subjective probability distributions.

Prioritization will play an important part in the assessment process. Prioritization will enable the determination of relative importance of an attribute and measure to mission effectiveness. The higher the priority, the higher the importance.

### **Frame of Reference - Scoring Models**

Aggregating the results of numerous measures to assess SUT operational effectiveness and suitability requires a common scale of “goodness” across the measures. A scorecard methodology that is based on a quantitative 0-to-1 scale or a discrete set of color codes (for example, red, yellow, green) provides a means to evaluate each measure and aggregate results into a normalized and presentable format.

---

<sup>8</sup> CJCSM 3500.04E, *Universal Joint Task Manual*, Dated August 25, 2008

<sup>9</sup> “Decision theory” is a branch of statistical theory concerned with quantifying the process of making choices between alternatives (Merriam-Webster Online Dictionary, <http://www.merriam-webster.com/dictionary/>, March 29, 2011)

Various scoring models may be constructed for this purpose, as long as each measure is based on a common scale and used consistently for every measure. We recommend using the JCIDS measures criteria of threshold and objective values as described earlier.

Threshold and objective values provide critical points for developing scores. System attributes (KPPs and KSAs) are typically evaluated as pass-fail based on meeting their threshold value. A simple scoring model that quantifies this pass-fail ordinal scale may exist as:

$$\begin{aligned} \text{measure value} < \text{threshold value, then score} &= 0 \\ \text{measure value} \geq \text{threshold value, then score} &= 1 \end{aligned}$$

Annex D provides descriptions of this and other models. When selecting a model: (1) simple is usually better; (2) stick with one model, using different models will skew results; and (3) model selection is ultimately up to the user.

### **Measurement Error**

T&E assesses operational effectiveness and suitability so that a decision-maker can determine whether to field the SUT. Thus, T&E must sufficiently reduce uncertainty and doubt (risk) about the SUT effectiveness and suitability to allow for an informed decision.

The assessment process must include an assessment on the level of risk or reliability of the information provided from the T&E process. This risk assessment is intended to quantify the possibility of making false conclusions about the SUT.

Errors are commonly known in hypothesis testing as type I and type II errors. If the risks of type I and type II errors can be quantified (estimated probability, cost, expected value, and so forth) then rational decision-making is improved. If a null hypothesis is incorrectly rejected when it is in fact true, this is called a type I error (also known as a false positive). Plainly speaking, it occurs in observing a difference when in truth there is none. A type II error (also known as a false negative) occurs when a null hypothesis is not rejected despite being false. This is the error of failing to observe a difference when in truth there is one. The Greek letter  $\alpha$  is used to denote the probability of type I error, and the letter  $\beta$  is used to denote the probability of type II error. Testers often tend to guard against the  $\beta$  error due to the possibility of fielding a deficient system.

### **SCOPING THE ASSESSMENT MODEL**

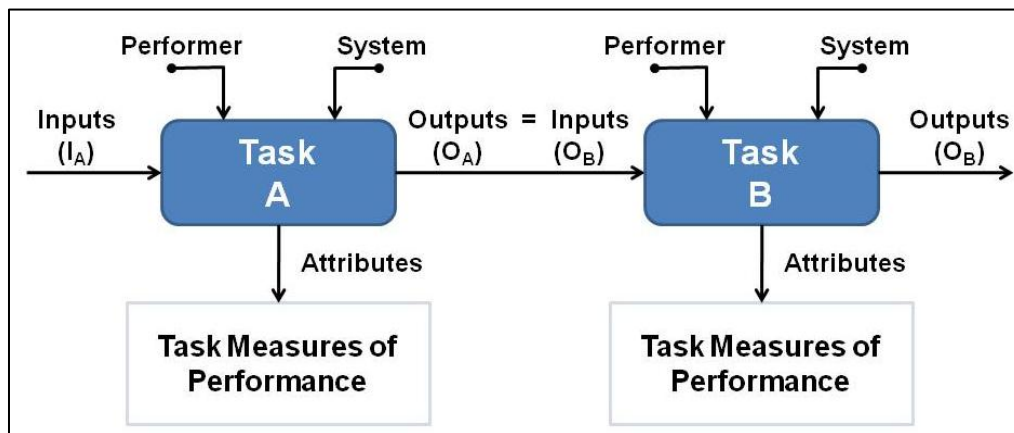
Testers often experience difficulty in identifying the boundaries of an assessment model. This section will offer some suggestions on how to scope the test in a way that supports relevant assessment of the SUT and its SoS. Once the appropriate mission is identified, relevant tasks and sub-tasks can be included in the test and evaluation process. This discussion is focused on the tasks and sub-tasks that must be evaluated in the test to assess the SUT impact on the SoS, task performance, and mission effectiveness.

### **Element Descriptions**

A task is defined as an action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability. Note that tasks are performed by individuals or organizations, not systems. A mission is composed of a set of tasks



that are conducted in series and parallel. In a simple case, a system performs tasks in serial (one at a time). Figure 1-3 illustrates a simple task model with one performer (individual or organization), one system, an input, and an output. The input may come from previous tasks as information, change in state, or change in resources. Note the output from task A then becomes the input for task B. Since these tasks are in series, the performer and the system could be the same for both tasks.



**Figure 1-3. Simple Task Model**

Many missions require more than one system to perform the tasks. Typically, this means that tasks are being performed in parallel supported by several systems as a SoS. The question then is to determine which tasks are relevant to the T&E of the SUT. It may be argued that any task (sub-task) in a mission thread are all equally important as one task impacts the remaining tasks down the thread. However, it may not be necessary or desirable to assess every task and sub-task, depending on the extent of the test.

This can be illustrated with the use of two examples. Figure 1-4 illustrates an example Operational Activity Decomposition Tree (DoDAF OV-5a). Similar to an integration definition (IDEF0) model, it shows a set of four tasks with inputs and outputs. Two systems perform the tasks with System<sub>2</sub> representing the SUT. Given this example, it may be necessary only to evaluate the three tasks performed by the SUT (that is, tasks 1, 2, and 4). However, the output from Task<sub>1</sub> acts as an input to Task<sub>3</sub> performed by System<sub>1</sub>. The output of Task<sub>3</sub> also acts as an input to the SUT performing Task<sub>4</sub>. These are SoS interactions that must be considered in the test design and task performance assessment process. If System<sub>1</sub> is available to be a part of the operational test, then it may be most efficient to evaluate Task<sub>3</sub> output. However, if System<sub>1</sub> is conceptually played in the operational test, then the output from Task<sub>1</sub> performed by the SUT needs to be evaluated to determine if it meets the requirements for System<sub>1</sub> to perform Task<sub>3</sub>. In this case, the output from Task<sub>3</sub> is assumed to exist for the SUT to use in Task<sub>4</sub>. The output of Task<sub>4</sub> then becomes the overall output of the set of tasks for the mission thread. Thus task performance measures must be developed not only for those tasks performed by the SUT, but also for those interactions (outputs) that it provides to other systems in the SoS and mission thread.



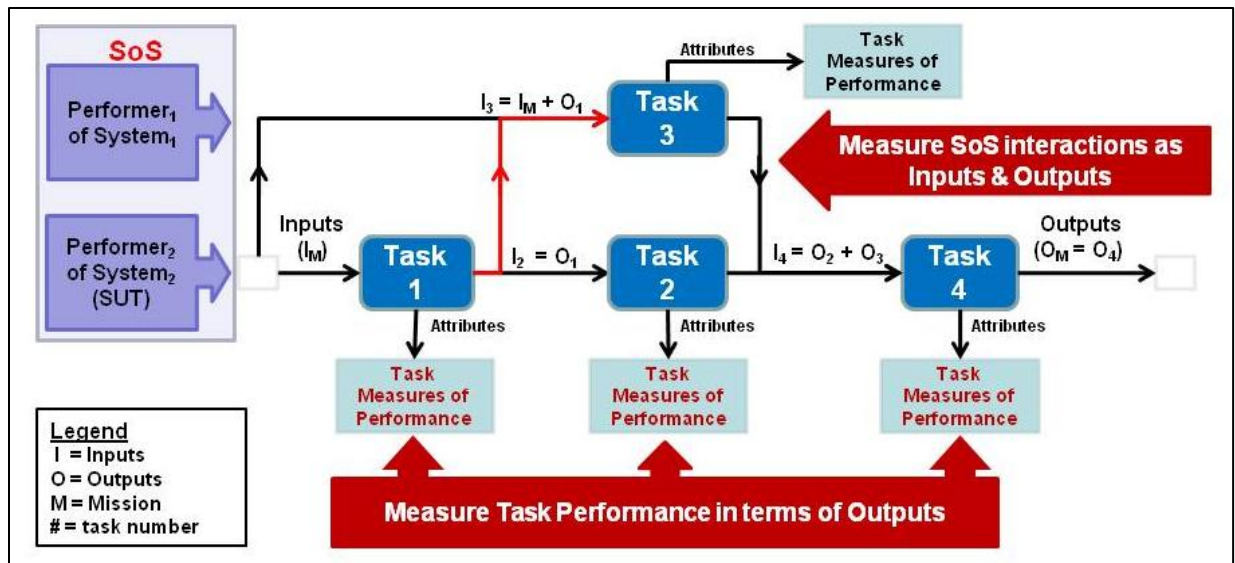


Figure 1-4. Complex Task Model (Example 1)

Suppose the OV-5a in figure 1-4 was altered by deleting the red arrow from Task<sub>1</sub> to Task<sub>3</sub>. In other words, Task<sub>1</sub> output did not interact with Task<sub>3</sub> as an input. In this case, there may not be a need to evaluate Task<sub>1</sub> separately. Task<sub>1</sub> and Task<sub>2</sub> could be combined and assessed as a single task with two sub-tasks. Figure 1-5 shows what this would look like. In this example, the SUT would require only an assessment of Task<sub>1</sub> and Task<sub>4</sub>. The inputs from Task<sub>3</sub> would be artificially inserted into the test.

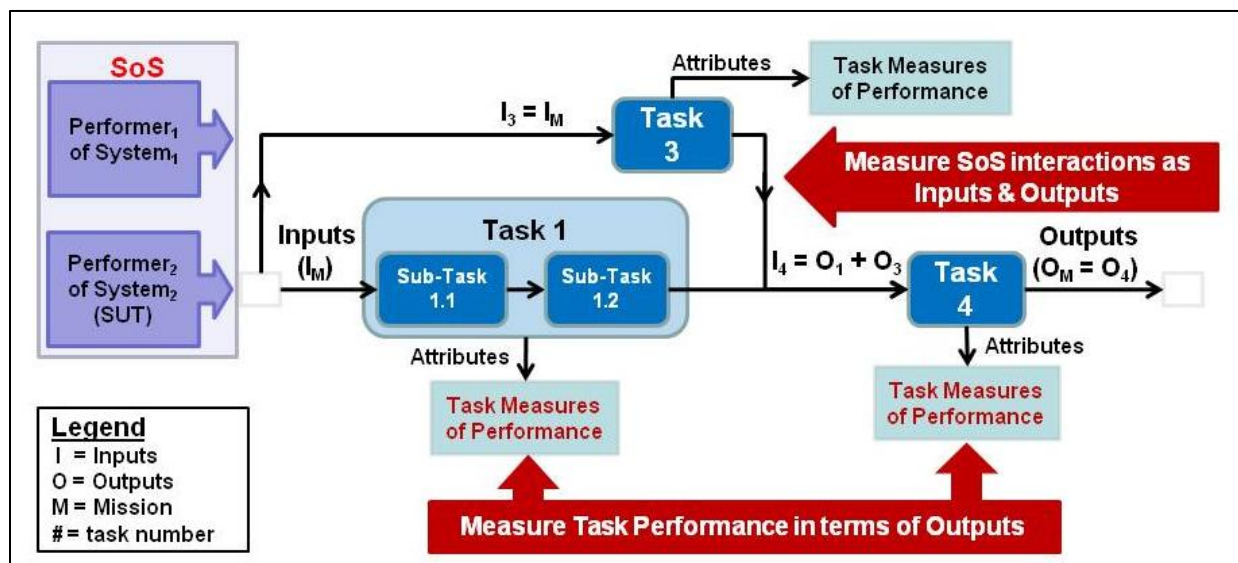


Figure 1-5. Complex Task Model (Example 2)

Tasks can be decomposed into sub-tasks. To determine what level of sub-tasks is sufficient to decompose the mission thread, apply the following basic rules:

1. A task has one primary performer and system associated with it. If there is more than one performer and system, then the task probably needs to be decomposed.
2. If the performer and system for the task provides an information exchange as an output during task performance (or other physical product), then the task should be decomposed so that the output is at the end of a sub-task.
3. Any outputs from a task performer that acts as an input to another task performer should occur at the end of a task (sub-task).
4. Decompose a task if it aids in assessing system functionality or in test design.

With some simple rules and analysis, the test design and analytical requirements of the test can be simplified to assess segments of a mission thread and still meet the need to evaluate the SUT as a component of a SoS conducting a mission thread.

## **GUIDEBOOK PLAN OF ACTION**

Chapter 2 of this guidebook provides the process of assessing a SUT's impact on task performance and mission effectiveness. Chapter 3 provides an example of that process using the JPR mission thread and a fictional PLB as the SUT. This process makes use of the measures decomposition process and example from the SOP to provide the framework for an assessment. To conduct this assessment, the following assumptions are made.

### **Assumptions**

The user of this guidebook:

- has developed mission, task, and system/SoS measures in accordance with the SOP
- has developed matrices 1 through 8 in the SOP
- is familiar with T&E practices and procedures to design and conduct tests
- is familiar with statistical analysis

## CHAPTER 2

### ASSESSMENT PROCESS

#### INITIATING THE ASSESSMENT PROCESS

The assessment process can begin once data has been collected in an integrated T&E process. Integrated T&E is intended to make use of all available and relevant data and information from contractor and government sources.<sup>10</sup> In most cases, early T&E will focus on system level attributes (KPPs, KSAs, and other attributes) and not on task and mission level attributes. The operational test team will address SUT impacts on task performance and mission effectiveness.

### STEP 1: INITIATING THE ASSESSMENT PROCESS

This step involves the process of verifying that relationships have been established across the measures framework and that components of the measures framework have been prioritized and weighted. Weighting allows the evaluator to pay proper attention to the missions in terms of the combat developer's priorities.

#### Relationship Mapping

##### Overview

This process verifies the relationships that have been established in the measures development phase. Those relationships are needed to:

- provide traceability of measure to attribute to its element (SUT, task, or mission)
- provide traceability of SUT impact on task performance and mission effectiveness

Given that measures were developed in accordance with the SOP, most of the necessary relationships will have already been established. Table 2-1 lists the relationships that should have been established and that play a vital role in the mission-based assessment process. Any gaps that may exist should be developed before proceeding to the assessment phase.

---

<sup>10</sup> Director, Operational Test and Evaluation Memorandum, Test and Evaluation Policy Revisions, December 22, 2007

**Table 2-1. Required Relationships for Mission-Based Assessment**

Level	Relationship	Rationale
Mission	Conditions to Mission	Needed to address conditions as specified in capability gaps
Mission	Desired Effects to Attributes	Needed to assess mission effects
Mission	Attributes to Measures	Needed to assess mission effects
Task	Tasks to Attributes	Needed to assess task performance
Task	Attributes to Measures	Needed to assess task performance
System/SoS	SUT to Attributes	Needed to assess SUT functionality
System/SoS	Attributes to Measures	Needed to assess SUT functionality
System - Task	System/SoS Attributes to Task Attributes	Needed to trace SUT/SoS impact on tasks
Task - Mission	Task Attributes to Mission Attributes	Needed to trace Task impact on mission effects

### Process

This process involves a verification and validation of relationship mapping for those elements in table 2-1. Table 2-2 illustrates an example matrix of task attributes mapped to mission attributes. This example matrix provides the traceability of task performance impact on mission desired effects through attributes at each level. Each matrix is intended simply to map relationships between the rows and columns.

**Table 2-2. Example Relationship Matrix (Task Attribute vs. Mission Attribute)**

Task - Attributes	Mission (Desired Effect) Attributes*					
	Mission Attribute 1	Mission Attribute 2	Mission Attribute 3	Mission Attribute 4	Mission Attribute 5	Mission Attribute 6
Task 1 – Attribute 1	X			X		
Task 1 – Attribute 2		X	X			
Task 2 – Attribute 1		X			X	
Task 3 – Attribute 1			X			
Task 3 – Attribute 2					X	X

\*NOTE: Mission attributes are based on attributes of mission desired effects. For simplicity, they will be referred to as “mission attributes” in this guidebook.

# Prioritization

## Overview

Mission effects, task performance, and SUT/SoS functions are all described by attributes. Prioritizing attributes enables the warfighter to express what is important. JCIDS lists prioritized attributes for the following four enabling Joint Capability Areas (JCA): battlespace awareness, command and control (C2), logistics, and net-centric (Senior Warfighters Forum attributes). These lists are a useful starting point for prioritizing attributes. JCIDS also details the process for developing system attributes in the form of KPPs and KSAs. KPPs are system attributes considered most critical or essential for an effective military capability. KSAs are system attributes considered critical or essential for an effective military capability, but not selected as KPPs. The important point is that KPPs will be of higher priority than KSAs when assessing system functions.

Measures are developed around attributes. Attributes are based on system/SoS functions, task performance criteria, and mission desired effects. Sometimes there will be more than one measure for each attribute; more than one attribute for each system, task, or mission; multiple tasks and missions; and several sets of conditions. This results in a complex design with which to assess system impacts. Prioritization provides a means to scope relative impacts on task performance and mission effects. Figure 2-1 illustrates the complexity of this issue and the need for prioritization.

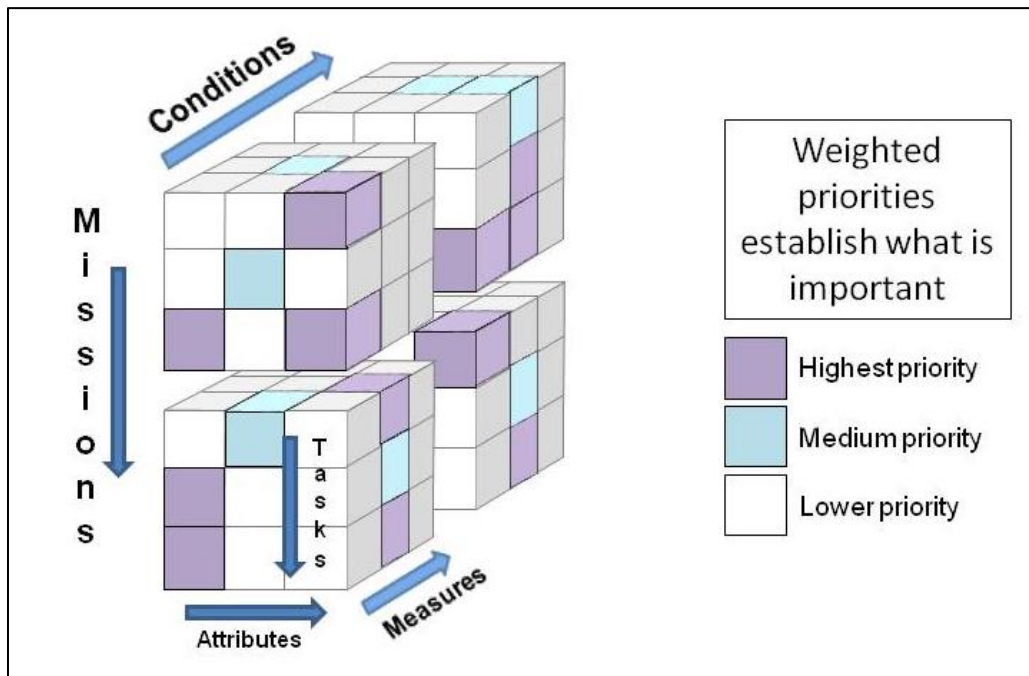


Figure 2-1. Prioritization Design Space

## Element Descriptions

- **Mission Prioritization**. A SUT that supports more than one mission needs to prioritize and weight each mission based on the relative importance of each mission and/or the probability of performing each mission. For example, if an aircraft is to be deployed conducting close air support 75% of the time and surveillance 25% of the time, the assessment should place greater emphasis on the close air support mission. Each  $W_M$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_M$  for each mission equal to 1. If missions are of equal importance, then the weighting will be the same.
  
- **Condition Prioritization**. A SUT may perform tasks and a mission under different sets of conditions to fill the capability gap identified in the capabilities-based assessment (CBA) document and initial capabilities document (ICD). Each set of conditions should be prioritized and weighted for a mission based on the relative importance of each condition set and/or the probability of performing each mission under that set of conditions. For example, if an aircraft is expected to be deployed conducting close air support in a desert type of terrain 75% of the time and a forested type of terrain 25% of the time, the assessment should place greater emphasis on the desert terrain conditions. Each  $W_c$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_c$  for each mission equal to 1.
  
- **Mission Effectiveness**. Mission effectiveness may be measured as a function of desired effects, mission attributes, and mission level measures (MLM).
  - **Mission desired effects** are used to assess mission effectiveness. For a single mission, there may be several desired effects with different priorities. Each desired effect should be weighted based on their relative importance to the warfighter. Each  $W_e$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_e$  for each mission equal to 1.
  
  - **Desired effect attributes** are used to assess the mission desired effect. For a single desired effect, there may be several attributes with different priorities. Each attribute should be weighted based on their relative importance to the warfighter. Each  $W_a$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_a$  for each mission equal to 1.

- **Mission measures** are used to assess the mission desired effect attribute. For a single attribute, there may be more than one measure with different priorities. Each measure should be weighted based on their relative ability to assess the attribute. Each  $W_m$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_m$  for each attribute equal to 1.
- **Task Performance**. Task performance may be measured as a function of the tasks performed, task attributes, and task level measures.
  - **Tasks and sub-tasks** for a mission must be evaluated to assess overall task performance. For a single mission, there may be a set of tasks and sub-tasks that are performed by the SUT and the SoS, each of which may need to be evaluated. Each task should be weighted based on their relative importance to the assessment of the SUT. Those tasks performed by the SUT should have the highest weight. Those tasks performed by other systems in the SoS but that rely on inputs from the SUT tasks may have lesser weight. Those tasks that are neither, but still a part of the mission may have zero weight. Each  $W_t$  shall be expressed as a rational number between 0 and 1.
  - **Task attributes** are used to assess the performance of a task. For each task, there may be more than one attribute with different priorities. Each attribute should be weighted based on their relative importance to the warfighter. Each  $W_a$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_a$  for each task equal to 1.
  - **Task measures** are used to assess the task attribute. For a single attribute, there may be more than one measure with different priorities. Each measure should be weighted based on their relative ability to assess the attribute. Each  $W_m$  shall be expressed as a rational number between 0 and 1 with the sum of  $W_m$  for each attribute equal to 1.
- **SUT/SoS Functionality**. SUT/SoS functionality may be measured as a function of the system/SoS attributes and the system/SoS level measures.
  - **System/SoS attributes** are used to assess the ability of a system and/or SoS to function as designed. Attributes are categorized as KPPs, KSAs, or other attributes. KPPs should have the highest weights, KSAs should be weighted less than KPPs, and other attributes should be weighted less than KPPs and KSAs. Each  $W_a$  should be expressed as a rational number between 0 and 1.

- **System/SoS measures** are used to assess the KPP, KSA, or other attribute. For a single attribute, there may be more than one measure with different priorities. Each measure should be weighted based on their relative ability to assess the attribute. Each  $W_m$  should be expressed as a rational number between 0 and 1 with the sum of  $W_m$  for each attribute equal to 1.

## Process

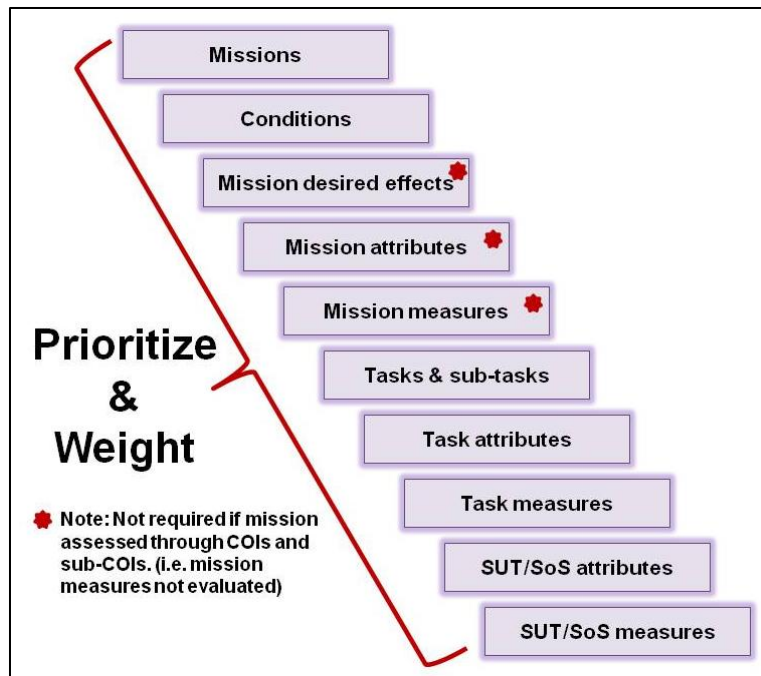
The process of prioritizing and weighting components of the measures framework will determine the significance measures have on assessing the SUT and its impact on the SoS, task performance, and mission effectiveness. Often, this process will utilize the matrixes developed in the measures decomposition process of the SOP to aid in the prioritizing and weighting activities.

### **Guidelines for Prioritizing and Weighting**

- Make use of all available authoritative sources to prioritize and weight components of the measures framework.
  - The capability documents (ICD, capability development document [CDD], and so forth) provides a good source for prioritization of missions, conditions, and tasks.
  - Use operations plans, concept of operations plans, and concept of operations to determine priorities.
- Mission essential tasks will be higher priority than supporting tasks.
- KPPs will have more weight than KSAs, and KSAs more than other attributes.
- The warfighter is the best source for determining priorities and weights.

The process involves determining priorities and weighting for each item shown in figure 2-2. It works best when starting at the mission level and working down to the system level.





**Figure 2-2. Prioritizing and Weighting Components**

### Process Shortcuts

This process is not as difficult as it appears, and there are shortcuts that can be taken to minimize the prioritization and weighting process:

- If the intent is to evaluate the mission and its desired effects subjectively through the use of critical operational issues (COI) and sub-COIs, then three components shown in figure 2-2 will not necessarily require weighting. Prioritization may still be completed for the desired effects simply to understand their importance in the reporting process. This shortcut in the evaluation process is discussed in the Measures Development SOP as an alternative approach when it is impractical to gather data on MLMs for a quantitative assessment.
- The easiest shortcut is to make all the weights equal as a starting point. For example, if there are two missions, weight each as 50%. If there are three desired effects for a mission, then weight each desired effect for that mission as 33.3%. This shortcut quickly populates the matrixes with weights to provide a starting point. As additional information is gathered that may add insight into priorities, these weights can be adjusted.
- Develop only one measure for each attribute. This will establish the weight as 100% for assessing that attribute. If a second measure is developed for an attribute, consider establishing a weighting schema for the primary measure and the secondary measure. For example, the primary measure may be weighted as 67% and the secondary measure weighted as 33%. If this is done, ensure it is documented and that it is done consistently.
- Prioritize and use weighting schemas for other components of the measures framework. For example, three attributes may be identified as relevant to evaluating a task. It may be desirable to prioritize these three attributes in order (1, 2, and 3). Follow a set weighting schema for the priorities. For example, priority 1 may be weighted 60%, priority 2 weighted 30%, and priority 3 weighted 10%.
- Place emphasis on collecting data for higher priority measures.

- Use commercially available decision support software (Decision Lens and so forth) to help in prioritization and weighting. These tools establish weighting based on consensus of subject matter experts (SME).

### Recording Priorities and Weights

The following describes the process for recording priorities and weights for the various components of the measures framework.

- **Mission and Condition Prioritization and Weighting.** A SUT that supports more than one mission needs to prioritize and weight each mission based on relative importance and/or the probability of performing the mission under a certain set of conditions. Table 2-3 provides an example format for recording this information.

**Table 2-3. Mission – Condition Weighting Example Format**

Mission Weighting ( $W_M$ )		Condition Weighting ( $W_C$ )			
Mission Weights $W_M$	Mission	Condition Set 1	Condition Set 2	Condition Set 3	Totals ↓
60%	Mission 1	75%	25%		100%
30%	Mission 2	50%	25%	25%	100%
10%	Mission 3		50%	50%	100%
100%	← Total				

The combined weight for a mission under a set of conditions can be determined as  $W_M * W_C$ . For example, the combined weight for mission 1 under condition set 1 is  $W_M = 0.6$  and  $W_C = 0.75$ ; therefore, the combined weight is  $0.6 * 0.75 = 0.45$ . This can be interpreted as meaning that 45% of the assessment will be focused on mission 1 under condition set 1.

- **Mission Effects Prioritization and Weighting.** For a single mission, assessing mission effects is a function of desired effects, attributes, and measures. The Measures Development SOP decomposes a mission using three separate matrices. (See figures C-1, C-2, and C-3 in annex C of this guidebook for examples.) Since these matrices map relationships, they can be used to determine priorities and develop weightings. Table 2-4 provides an example format for recording this information.

**Table 2-4. Mission Effects Weighting Example Format**

Matrix 1 ↓								
Mission	Desired Effect Weights ( $W_e$ )			Total	Mission Measure Weights ( $W_m$ )			
Mission 1	60%	25%	15%	100%	Measure 1	Measure 2	Measure 3	Totals
	Desired Effect 1	Desired Effect 2	Desired Effect 3	Attributes				↓
Attribute Weights ( $W_a$ )	50%		100%	1	50%	50%		100%
	25%	50%		2	75%		25%	100%
	25%	50%		3		100%		100%
	100%	100%	100%	← Totals				
Matrix 2 ↑					Matrix 3 ↑			

The weight that a single measure can have on a mission through a single attribute and single desired effect can be determined as  $W_m * W_a * W_e$ . For example, the combined weight for measure 1 on attribute 2 for desired effect 1 is  $W_m = 0.75$ ,  $W_a = 0.25$ , and  $W_e = 0.6$ ; therefore, the calculated weight is  $0.75 * 0.25 * 0.6 = 0.1125$ . This can be interpreted as meaning that approximately 11% of the assessment will be based on that combination of measure 1, attribute 2, and desired effect 1.

- **Task Performance Prioritization and Weighting.** For a single mission, assessing task performance is a function of the tasks, attributes, and measures. The Measures Development SOP decomposes a mission into tasks, attributes, and measures using three separate matrices. (See figures C-5, C-6, and C-7 in annex C of this guidebook for examples.) Since these matrices map relationships, they can be used to determine priorities and develop weightings. Table 2-5 provides an example format for recording this information.

**Table 2-5. Task Performance Weighting Example Format**

Matrix 4 ↓										
Mission	Task Weights ( $W_t$ )					Total	Measure Weights ( $W_m$ )			
Mission 1	25%	25%	12.5%	25%	12.5%	100%	Measure 1	Measure 2	Measure 3	Total ↓
SUT Performed Task	Y	Y	N	Y	N					
	Task 1	Task 2	Task 3	Task 4	Task 5	Attributes				
Attribute Weights ( $W_a$ )	50%		75%	100%		1	50%	50%		100%
	50%	50%			100%	2	75%		25%	100%
		50%	25%			3		100%		100%
	100%	100%	100%	100%	100%	← Totals				
Matrix 5 ↑						Matrix 6 ↑				

The process to weight the tasks for a mission is not included in the Measures Development SOP. However, it is covered in chapter one of this guidebook in the “Scoping” section. Tasks are categorized as either (1) performed by the SUT; (2) performed by other systems, but receive direct input from the SUT; or (3) performed by other systems without input from the SUT. It is suggested that a weighting schema is used that is based on the following three rules:

- Every task is equally important, as the mission cannot be completed without any one of the tasks.
- Tasks performed by other systems and with input from the SUT are of some importance to the T&E.
- Tasks performed by other systems and without input from the SUT may be of little importance to the T&E.

Based on the scoping and established rules, the tasks can be weighted. Table 2-6 illustrates how the five tasks in table 2-5 would be weighted based on a simple weighting schema. The calculated weights in column 5 are used for the task weights in table 2-5.

**Table 2-6. Task Weighting Schema**

Column 1	Column 2	Column 3	Column 4	Column 5
Task Category	Count of Tasks in Category	Schema Weight	Total Weight (Col 2 * Col 3)	Calculated Weight for Each Task (Col 3/Sum)
1. Task performed by SUT	3	2.0	6.0	0.25
2. Task performed by other systems but relies on input from SUT	2	1.0	2.0	0.125
3. Task performed by other systems	0	0.0	0.0	0.0
Sum			8.0	

The weight that a single measure can have on a mission through a single task and a single attribute can be determined as  $W_m * W_a * W_t$ . For example, the combined weight for measure 1 on attribute 2 for task 1 is  $W_m = 0.75$ ,  $W_a = 0.50$ , and  $W_t = 0.25$ ; therefore, the calculated weight is  $0.75 * 0.50 * 0.25 = 0.09375$ . This can be interpreted as meaning that approximately 9% of the assessment will be based on that combination of measure 1, attribute 2, and task 1.

- **System/SoS Attribute Prioritization and Weighting.** A SUT is assessed in its ability to function with specified characteristics and technical attributes. The Measures Development SOP recognizes KPPs, KSAs, and other attributes as part of the evaluation for effectiveness and suitability. (See figures C-8, C-9, and C-10 in annex C of this guidebook for examples from the SOP.) Since these matrices map relationships, they can be used to help determine priorities and develop weightings. Table 2-7 provides an example format for recording this information.

**Table 2-7. System/SoS Attribute Weighting Example Format**

System/SoS Attribute			Measure		Condition Weights ( $W_c$ )			
Type	Weight ( $W_a$ )	Attribute	Weight ( $W_m$ )	Measure	Condition Set 1	Condition Set 2	Condition Set 3	Totals
KPP	25.0%	1	60%	Measure 1	60%	27.5%	12.5%	100%
			40%	Measure 2				
KPP	25.0%	2	100%	Measure 3	60%	27.5%	12.5%	100%
KSA	16.6%	3	80%	Measure 4	100%			100%
			20%	Measure 5				
KSA	16.6%	4	100%	Measure 6	60%	27.5%	12.5%	100%
OA*	8.4%	5	100%	Measure 7	100%			100%
OA	8.4%	6	100%	Measure 8	100%			100%
	100%	Totals						

\*OA (Other Attribute)

The attribute weights in table 2-7 are based on a weighting schema as illustrated in table 2-8. KPPs have the highest weights, KSAs have the second level of weights, and other attributes have the lowest weights. For simplicity, this schema assumes all attributes in the same category are equal in priority (that is, all KPPs are equal and so forth). The calculated weights in column 5 are used for the attribute weights in table 2-7.

**Table 2-8. System/SoS Attribute Weighting Schema**

Column 1	Column 2	Column 3	Column 4	Column 5
Attribute Category	Count of Attributes in Category	Schema Weight	Total Weight (Col 2 * Col 3)	Calculated Weight for Each Attribute (Col 3/Sum)
KPP	2	3.0	6.0	0.25
KSA	2	2.0	4.0	0.166
Other Attribute	2	1.0	2.0	0.084
<b>Sum</b>			12.0	

Since system attributes may be impacted by conditions, the condition sets are added to the system/SoS attribute weighting matrix. Typically, developmental testing will evaluate system attributes across various descriptors for conditions to ensure the system functions according to technical specifications. The Operational Tester may continue to test system attributes across mission condition sets. Since a condition set may apply across several missions, the condition weighting on system/SoS attributes must be calculated based on the mission and condition set weightings from table 2-3. The formula for calculating the weights is:

For example, the calculated weight for condition set 1 across all three missions is  $(0.75 * 0.60) + (0.50 * 0.30) + (0.0 * 0.10) = 0.45 + 0.15 + 0.0 = 0.60$ . This is the weight placed under condition set 1 in figure 2-6. Note that there are several attributes where the weight values for condition set 1 are 100%. In this example, condition set 1 is considered the “base” condition set in which those attributes will only be evaluated under that condition set. This is a practice that may help to simplify the assessment of system/SoS attributes.

## STEP 2: CONDUCT THE ASSESSMENT

This step assumes that the test has been completed and that adequate data has been collected on the system/SoS, task, and mission measures. The assessment is a bottom-up approach in that the system/SoS attributes are assessed first, then the task is performed, and finally the mission level is assessed. This is the point where ascending the right side of the “V” diagram in figure 1-1 begins.

### Element Descriptions

- **Single Measure Scoring Model.** Measure scoring models are discussed in chapter 1 with some example models illustrated in annex D. For the purposed of this assessment, the threshold model will be used. The model results in a score of 0 or 1 based on whether the measure meets the threshold.
- **Aggregate Measure Scoring Model.** The aggregate measure scoring model takes measures that are aggregated into one numerical score and color codes them to provide a visual interpretation of the value. Table 2-9 illustrates a simple five-level aggregate measure scoring model that color codes score values. The colors provide a visualization of the assessment for individual and aggregated measure scores.

Table 2-9. Measure Color Codes

Color Code	Aggregate Measure Score
Green	0.81 to 1.00
Light Green	0.61 to 0.80
Yellow	0.41 to 0.60
Orange	0.21 to 0.40
Red	0.00 to 0.20

- **Risk Model.** Chapter 1 discusses the possibility of measurement error. This may be put in terms of the “risk of drawing a wrong conclusion” by (1) concluding that the measure satisfied the threshold value when it did not or (2) concluding that the measure did not satisfy the threshold value when it did. There are numerous textbooks on design of experiments and statistics that address confidence levels to minimize risk. However, real life resource and time constraints do not always allow for sufficient data to be collected. A risk model is suggested to evaluate the level of risk in making conclusions that is based on the available data. If the risk level is too high, then the evaluation team may ask for additional testing to provide sufficient data to add confidence to the test results.

The risk model presented here is based on a commonly accepted risk matrix as shown on the left side of figure 2-3. Risk is evaluated in terms of the “likelihood of error in drawing a conclusion” and the “level of impact in drawing the wrong conclusion.” The risk matrix uses a 1 to 5 scale for both likelihood and impact. The lower the number for likelihood of error,

the least likely error will occur (higher confidence). The lower the number for impact level, the less impact the error will have on the assessment.

Likelihood level	5					
	4					
	3					
	2					
	1					
		1	2	3	4	5
Impact level						

Likelihood level	5	0.20	0.40	0.60	0.80	1.00
	4	0.16	0.32	0.48	0.64	0.80
	3	0.12	0.24	0.36	0.48	0.60
	2	0.08	0.16	0.24	0.32	0.40
	1	0.04	0.08	0.12	0.16	0.20
		1	2	3	4	5
Impact level						

**Figure 2-3. Risk Matrix (Left Side – Standard Matrix; Right Side – Adjusted Matrix)**

The matrix on the right side of figure 2-3 calculates values for each square of the matrix (that is, calculated as Row \* Column/maximum score of 25). This matrix is used to construct a common set of color-coded scores as illustrated in table 2-10. This table will be used for color coding risk in the assessment process.

**Table 2-10. Risk Color Codes**

Color	Risk Factor Levels
Green	0.00 to 0.31
Yellow	0.32 to 0.59
Red	0.60 to 1.00

- o **Likelihood Levels.** Risk includes the likelihood of occurrence. With respect to measures, likelihood of occurrence refers to the likelihood or probability that the measure provides a correct assessment in terms of reliability and validity. Reliability describes the repeatability and consistency of a test. Validity defines the strength of the final results and whether they can be regarded as accurately describing the real world. A measure's reliability and validity may be based on statistical inferences and the ability to minimize bias in the test. It is the ability to minimize type I and type II errors and make an incorrect conclusion about cause and effect. Table 2-11 defines levels of reliability and validity that can then be used in determining the likelihood of occurrence.



**Table 2-11. Levels of Likelihood - Reliability and Validity**

No.	Level	Definition
1	Significant	Measure is statically significant in terms of reliability and validity
2	High	Measure is statically significant in terms of validity, but not reliability
3	Medium	Measure is not statically significant due to lack of sufficient data, but is assessed as reliable and valid (low variance and meets expected results)
4	Low	Measure assessed as valid, but not reliable (high variance)
5	Null	Measure assessed as neither valid nor reliable

- o **Impact Levels.** Impacts differ for each level of assessment.
  - **Task Level Measure and MLM Impact.** Task level measure and MLM impact are based on the weights determined in step 1. For example, a task level measure will be based on the weight for the measure, times the weight of the attribute, times the weight of the task (that is,  $\text{Impact} = W_t * W_a * W_m$ ).
  - **System/SoS Attribute Impact Levels.** Impact levels of the system/SoS attributes will be based on the type of attribute. Table 2-12 illustrates an example method for quantifying impact levels for system/SoS attributes.

**Table 2-12. Impact Levels**

Attribute Type	Impact Levels
KPP	5
KSA	3
Other Attribute	1

- **Exceptions Reporting.** Exceptions reports provides a visual listing of SUT and SoS strengths, weaknesses, constraints, limitations, and other issues that may be considered important to highlight for the decision-making authority. These are items that should help to explain deficiencies in the reporting of system/SoS functionality, task performance, and mission effectiveness.

## System/SoS Assessment

### Overview

The purpose of the system/SoS evaluation is to determine the functional capabilities of the system/SoS when employed in a realistic operational environment. This involves the assessment of system/SoS attributes for both effectiveness and suitability.

## Process

The process includes an evaluation of data to determine scores for each measure and the measure reliability (risk), consolidate the measures and attributes, and identify exceptions.

- **Evaluation of Data.** The evaluation process is relatively straightforward because the standards needed for evaluation should have been developed in a test plan. The process, regardless of the testing source (developmental or operational), begins by comparing test results with established standards.
- **Scoring Measures.** Scoring individual measures involves comparing the data with the established threshold and objective values. If not already done, a scoring model must be selected from annex D of this guidebook or one must be developed by the assessment team. For discussion purposes, the Threshold Scoring Model is selected for use in this guidebook. This will result in scores of 0 or 1 based on meeting the threshold value. Every measure for each effectiveness and suitability attribute of the system/SoS should be scored based on the selected model. The results can be summarized in tabular format as illustrated in the example table 2-8. Although the scoring occurs at the measure level, the results can be aggregated to the attribute level by simple weighting of each measure for that attribute. Columns 3 and 4 in table 2-13 show the weight and measure for each attribute. Recall that system/SoS attributes in table 2-7 may need to be tested under different sets of conditions. Those condition sets are weighted for the attribute and measures (shown in table 2-13 as columns 7 and 8). Note that the measure threshold and objective values may differ across condition sets. Each row in column 10 of table 2-13 shows the score values for a single measure in a single condition set. These are aggregated for a single attribute as one value in column 11 based on weighted sums for the scores across measures and condition sets. This can be calculated as:

For example, the total score for attribute one in table 2-13 is the sum of the six rows of data for the two measures and three condition sets, that is, the Sum of (Col 3)(Col 8)(Col 10). The calculated value is 0.7.

Note that although table 2-13 is for effectiveness attributes, a similar table can be constructed for suitability attributes. The score for each attribute can then be color coded based on table 2-9. In this example, we see that attribute 2 (a KPP) is scored low and would not pass the test. Attribute 1 may also not pass since it did not meet all of the measure threshold values.

**Table 2-13. System/SoS Scoring Table**

Effectiveness										
Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11
Attribute	Type	Measure Weight (W <sub>m</sub> )	Measure	Threshold value (TV)	Objective value (OV)	Condition Set	Condition Weight (W <sub>c</sub> )	Observed value	Score value (S <sub>cm</sub> )	Total Score (S <sub>a</sub> )
1	KPP	60%	Measure 1	TV1-1	OV1-1	1	60%	> TV1-1	1	0.915
				TV1-2	OV1-2	2	27.5%	> TV1-2	1	
				TV1-3	OV1-3	3	12.5%	< TV1-3	0	
		40%	Measure 2	TV2-1	OV2-1	1	60%	> TV2-1	1	
				TV2-2	OV2-2	2	27.5%	< TV2-2	0	
				TV2-3	OV2-3	3	12.5%	> TV2-3	1	
2	KPP	100%	Measure 3	TV3-1	OV3-1	1	60%	< TV3-1	0	0.4
				TV3-2	OV3-2	2	27.5%	> TV3-2	1	
				TV3-3	OV3-3	3	12.5%	> TV3-3	1	
3	KSA	80%	Measure 4	TV4-1	OV4-1	1	100%	> TV4-1	1	0.8
		20%	Measure 5	TV5-1	OV5-1			< TV5-1	0	
4	KSA	100%	Measure 6	TV6-1	OV6-1	1	60%	> TV6-1	1	1
				TV6-2	OV6-2	2	27.5%	> TV6-2	1	
				TV6-3	OV6-3	3	12.5%	> TV6-2	1	
5	OA	100%	Measure 7	TV7-1	OV7-1	1	100%	> TV7-1	1	1
6	OA	100%	Measure 8	TV8-1	OV8-1	1	100%	> TV8-1	1	1

- Measuring Risk.** Various parametric and non-parametric statistical tests may be applied to the data in order to manage risk by improving the reliability of the data. If it can be statistically determined that the observed value is better than the threshold value, then there is probably a low risk at scoring the measure as satisfactory. However, if there is insufficient data to conduct any statistical tests, then the risk of making a conclusive determination is increased. If the observed value is close to the threshold value, then there is also a higher risk of error. Table 2-11 provided some guidelines for scoring risk in terms of reliability and validity. The assessment team may wish to establish additional guidelines for determining risk.

Risk is also based on impact. Table 2-12 provided impact values based on the type of attribute. The risk assessment for the system/SoS attributes can be added to table 2-13 as two new columns. The new consolidated table is shown as table 2-14. Note the column numbers correspond to the columns in table 2-13. Four new columns are added (column 12 through column15) to determine risk. Column 12 comes from table 2-12 for each attribute. Column 13 comes from table 2-11 based on the data and evaluation of the data for each measure and condition set. A risk score for each attribute (R<sub>a</sub>) is calculated using the following formula:

And risk score ( $R_a$ ) is the calculated risk divided by the maximum score (25) from figure 2-3:

Based on table 2-14, this calculation is the sum of (Col 3)\* (Col 8)\* (Col 12)\* (Col 13).

**Table 2-14. System/SoS Risk Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 7	Col 8	Col 11	Col 12	Col 13	Col 14	Col 15
Attribute	Type	Measure Weight ( $W_m$ )	Measure	Condition Set	Condition Weight ( $W_c$ )	Total Score ( $S_a$ )	Impact ( $I_a$ )	Likelihood of error ( $L_{cm}$ )	Calc. Risk	Risk Score ( $R_a$ )
1	KPP	60%	Measure 1	1	60%	0.915	5	1	6.8	0.272
				2	27.5%			1		
				3	12.5%			3		
		40%	Measure 2	1	60%			1		
				2	27.5%			2		
				3	12.5%			3		
2	KPP	100%	Measure 3	1	60%	0.4	5	1	9.625	0.385
				2	27.5%			3		
				3	12.5%			4		
3	KSA	80%	Measure 4	1	100%	0.8	3	1	3.6	0.144
		20%	Measure 5					2		
4	KSA	100%	Measure 6	1	60%	1	3	2	7.2	0.288
				2	27.5%			3		
				3	12.5%			3		
5	OA	100%	Measure 7	1	100%	1	1	1	1	0.04
6	OA	100%	Measure 8	1	100%	1	1	2	2	0.08

Referring back to table 2-10, the risk for each attribute ( $R_a$ ) would be green, except for attribute 2, which would be yellow. This gives the decision-maker confidence in making an assessment for each of the attributes as to whether they meet the threshold values.

- Aggregating Scores.** Aggregated scores for system/SoS attributes may be determined for both attribute measures and attribute risk. They are both based on the weights assigned to the types of attributes found in table 2-8. Aggregating scores allows the decision-maker to see a single value that can be used as a gauge for SUT functional effectiveness and suitability. Note that it may be desirable to have one value for functional effectiveness and one value for functional suitability.
  - Aggregate Attribute Measures.** Tables 2-8 and 2-13 provide the information needed to aggregate attribute measures into one score. The process simply uses values in column 5 of table 2-8 as weights for the scores in column 11 of table 2-13. The combined information and calculated aggregate score is shown in table 2-15 using the calculation for a SUT attribute effectiveness of:

**Table 2-15. Aggregate System/SoS Scoring Table**

Table 2-8	Table 2-13			New Column
Column 5	Column 1	Column 2	Column 11	
Calc. Attribute Weight ( $W_a$ )	Attribute	Type	Score ( $S_a$ )	Aggregate Score ( $S_{eff}$ )
0.25	1	KPP	0.915	0.796
0.25	2	KPP	0.4	
0.166	3	KSA	0.8	
0.166	4	KSA	1	
0.084	5	OA	1	
0.084	6	OA	1	

Based on the color codes for scores in table 2-9, the aggregated score for SUT effectiveness attributes would be colored light green to indicate an almost 80% weighted score.

- Aggregate Attribute Risk.** The process for aggregating attribute risk for the system/SoS is similar to the process used to aggregate attribute measures. Tables 2-8 and 2-14 provide the information needed to aggregate attribute risk into one score. The process simply uses values in column 5 of table 2-8 as weights for the attributes and the risk score in column 15 of table 2-14. The combined information and calculated aggregate score is shown in table 2-16 using the calculation for a SUT attribute effectiveness of:

**Table 2-16. Aggregate System/SoS Risk Scoring Table**

Table 2-8	Table 2-14			New Column
Column 5	Column 1	Column 2	Column 15	
Calc. Attribute Weight ( $W_a$ )	Attribute	Type	Risk Score ( $R_a$ )	Aggregate Risk Score ( $R_{eff}$ )
0.25	1	KPP	0.272	0.246
0.25	2	KPP	0.385	
0.166	3	KSA	0.144	
0.166	4	KSA	0.288	
0.084	5	OA	0.04	
0.084	6	OA	0.08	

Based on the color codes for scores in table 2-10, the aggregated risk score for SUT effectiveness attributes would be colored green to indicate minimal risk in the SUT effectiveness attribute scores.

- Identification of Exceptions.** An exceptions report should be included with the SUT attributes scores and risk scores to explain and highlight deviations in the scores. It provides a means to show details on areas of concern. A sample format for an exceptions report is illustrated as table 2-17.

**Table 2-17. System/SoS Attribute Exceptions**

Column 1	Column 2	Column 3	Column 4
<b>Effectiveness Attributes</b>			
ID No.	Attribute	Type	Exception
E1	1	KPP	Issue
E2	3	KSA	Issue
E3	4	KSA	Issue
<b>Suitability Attributes</b>			
ID No.	Attribute	Type	Exception
S1	7	KPP	Issue
S2	8	KSA	Issue
S3	9	OA	Issue

## Task Performance Assessment

### Overview

The purpose of the task performance assessment is to determine whether the SUT provides the necessary capability to support the SoS in the performance of tasks. This begins to place the focus on the needs of the warfighter to perform tasks in order to achieve mission desired effects. The assessment is still focused on the SUT, but now in terms of how the warfighter can perform their tasks with the SUT as a part of the SoS in an operationally realistic environment.

This process requires that task measures have established threshold values and, possibly, objective values. Historically, these values have not been included in the capability development document (CDD), but may reside in the CBA, ICD, CONOPS, OPLANs, or other doctrine. In a worst case scenario, the warfighter should be able to provide threshold values for task measures.

### Process

The process to assess system/SoS attributes is similar to that in the previous section. What differs is that the assessment examines attributes of the tasks and not attributes of the SUT. This process will include evaluating data, determining scores for each measure and the measure reliability (risk), consolidating the measures and attributes, and identifying exceptions.

- **Evaluation of Data.** The process begins by comparing test results with established standards.
- **Scoring Measures.** Scoring individual measures involves comparing the data with the established threshold and objective values. The scoring model used should be the same as that selected for the system/SoS attributes. However, an alternate scoring model may be selected, but must be consistently used for all task assessments. For discussion purposes, the Threshold Scoring Model will continue to be used in this guidebook. This will result in scores of 0 or 1 based on meeting the threshold value.

Recall that task measures are developed based on performance attributes of the tasks (refer to the Measures Development SOP), and that each task, attribute, and measure is weighted. This will provide the framework for assessing task performance. Every measure should be scored based on the selected scoring model. The results can be summarized in tabular format as illustrated in the example table 2-18. This example draws from the Mission 1 tasks and weights shown in table 2-5 and the two conditions sets for Mission 1 shown in table 2-3. Although the scoring occurs at the measure level, they can be aggregated up to the task level by simple weighting of each measure and attribute.

**Table 2-18. Task Performance Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11	Col 12
Task	Condition Weight (W <sub>c</sub> )	Condition Set	Attribute Weight (W <sub>a</sub> )	Attribute	Measure Weight (W <sub>m</sub> )	Measure	Threshold value (TV)	Objective value (OV)	Observed value (Obv)	Score value (S <sub>cam</sub> )	Total Score (S <sub>t</sub> )
1	75%	1	50%	1	100%	1	TV1-1	OV1-1	> TV1-1	1	0.81
			50%	2	75%	1	TV1-2	OV1-2	> TV1-2	1	
			25%	3	25%	3	TV3-1	OV3-1	< TV3-1	0	
	25%	2	50%	1	100%	1	TV1-3	OV1-3	> TV1-3	1	
			50%	2	75%	1	TV1-4	OV1-4	< TV1-4	0	
			25%	3	25%	3	TV3-2	OV3-2	> TV3-2	1	
2	75%	1	50%	2	75%	1	TV1-5	OV1-5	> TV1-5	1	0.44
			25%	3	25%	3	TV3-3	OV3-3	< TV3-3	0	
			50%	3	100%	2	TV2-1	OV2-1	< OV2-1	0	
	25%	2	50%	2	75%	1	TV1-6	OV1-6	< TV1-6	0	
			25%	3	25%	3	TV3-4	OV3-4	> TV3-4	1	
			50%	3	100%	2	TV2-2	OV2-2	> OV2-2	1	
3	75%	1	75%	1	100%	1	TV1-7	OV1-7	> TV1-7	1	0.75
			25%	3	100%	2	TV2-3	OV2-3	< OV2-3	0	
	25%	2	75%	1	100%	1	TV1-8	OV1-8	> TV1-8	1	
			25%	3	100%	2	TV2-4	OV2-4	> OV2-4	0	
4	75%	1	100%	1	100%	1	TV1-9	OV1-9	> TV1-9	1	1.00
	25%	2	100%	1	100%	1	TV1-10	OV1-10	> TV1-10	1	
5	75%	1	100%	2	75%	1	TV1-11	OV1-11	> TV1-11	1	0.81
					25%	3	TV3-5	OV3-5	< TV3-5	0	
	25%	2	100%	2	75%	1	TV1-12	OV1-12	> TV1-12	1	
					25%	3	TV3-6	OV3-6	> TV3-6	1	

Arbitrary observed values are assigned in column 10 for illustration purposes. These values would come from the evaluation of data. The score is then assigned in column 11 based on the scoring model. The total score for each task is then calculated based in the weights for the conditions, attributes, and measures and the scores in column 11. This can be calculated as:

For example, the total score for task 1 in the table is the sum of the six rows of data for the two condition sets, two attributes, and two measures; that is, the sum of (Col 2)\* (Col 4)\* (Col 6)\* (Col 11). The calculated value is 0.81. The results shown in column 12 of table 2-12 can be color coded based on the color code shown in table 2-9. In this example, tasks 1, 4, and 5 are green; task 3 is colored light green; and task 2 is colored yellow.

- **Measuring Risk.** Just as risk was determined for the system/SoS attributes in the previous section, risk can also be determined for the task performance scores. Table 2-11 provides guidelines for scoring risk in terms of reliability and validity. The assessment team may wish to establish additional guidelines for determining risk. Risk is also based on impact. Impact is the weight place on each of the conditions, attributes, and measures. The risk assessment for task performance scores can be added to table 2-18 as three new columns. The new consolidated table is shown as table 2-19. Note the column numbers correspond to the columns in table 2-18. Three new columns (13 through 15) are added to determine risk. Column 13 comes from table 2-11 based on the data and evaluation of the data for each measure and condition set. For illustration purposes, the values in column 13 are arbitrary. A risk score for each task ( $R_t$ ) is calculated using the following formula:

Risk score ( $R_t$ ) is the calculated risk divided by the maximum score (25) from figure 2-3:

---

As illustrated in table 2-19, this calculation is the sum of (Col 2)\* (Col 4)\* (Col 6)\* (Col 13).

Referring back to table 2-10, the risk for each attribute ( $R_t$ ) would be green. This gives the decision-maker confidence in making an assessment for each of the attributes as to whether they meet the threshold values.



**Table 2-19. Task Performance Risk Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 12	Col 13	Col 14	Col 15
Task	Condition Weight (W <sub>c</sub> )	Condition Set	Attribute Weight (W <sub>a</sub> )	Attribute	Measure Weight (W <sub>m</sub> )	Measure	Total Score (S <sub>t</sub> )	Likelihood of error (L <sub>cam</sub> )	Calc. Risk	Risk Score (R <sub>t</sub> )
Task 1	75%	1	50%	1	100%	1	0.81	1	1.25	0.05
			50%	2	75%	1		1		
			50%	2	25%	3		3		
	25%	2	50%	1	100%	1		1		
			50%	2	75%	1		1		
			50%	2	25%	3		3		
Task 2	75%	1	50%	2	75%	1	0.44	1	1.75	0.07
			50%	3	100%	2		3		
			50%	3	100%	2		2		
	25%	2	50%	2	75%	1		1		
			50%	3	100%	2		3		
			50%	3	100%	2		2		
Task 3	75%	1	75%	1	100%	1	0.75	1	1.25	0.05
			25%	3	100%	2		2		
	25%	2	75%	1	100%	1		1		
			25%	3	100%	2		2		
Task 4	75%	1	100%	1	100%	1	1.00	1	1.0	0.04
	25%	2	100%	1	100%	1		1		
Task 5	75%	1	100%	2	75%	1	0.81	1	1.5	0.06
					25%	3		3		
	25%	2	100%	2	75%	1		1		
					25%	3		3		

- **Aggregating Scores.** Aggregated scores for task performance attributes may be determined for both measures and risk. They are based on the weights assigned to the tasks found in table 2-6. Aggregating scores allows the decision-maker to see a single value that can be used as a gauge for SUT impact on task performance.
  - **Aggregate Task Performance Measures.** Tables 2-6 and 2-18 provide the information needed to aggregate attribute measures into one score. The process simply uses values in column 5 of table 2-6 as weights for the scores in column 12 of table 2-18. The combined information and calculated aggregate score is shown in table 2-20 using the calculation of:

**Table 2-20. Aggregate Task Performance Scoring Table**

Table 2-6	Table 2-18		New Column
Column 5	Column 1	Column 12	
Calc. Task Weight ( $W_t$ )	Task	Score ( $S_t$ )	Aggregate Score ( $S_{task}$ )
0.25	1	0.81	<b>0.758</b>
0.25	2	0.44	
0.125	3	0.75	
0.25	4	1.00	
0.125	5	0.81	

Based on the color codes for scores in table 2-9, the aggregated score for task performance would be colored light green indicating an almost 76% weighted score.

- o **Aggregate Task Performance Risk.** The process for aggregating task performance risk in the assessment again uses information from tables 2-6 and 2-19. The process simply uses values in column 5 of table 2-6 as weights for the task and the risk score in column 15 of table 2-19. The combined information and calculated aggregate score is shown in table 2-21 using the calculation :

**Table 2-21. Aggregate Task Performance Risk Scoring Table**

Table 2-6	Table 2-19		New Column
Column 5	Column 1	Column 15	
Calc. Task Weight ( $W_a$ )	Task	Risk Score ( $R_a$ )	Aggregate Risk Score ( $R_{eff}$ )
0.25	1	0.05	<b>0.054</b>
0.25	2	0.07	
0.125	3	0.05	
0.25	4	0.04	
0.125	5	0.06	

Based on the color codes for scores in table 2-10, the aggregated risk score for task performance would be green, indicating minimal risk in the assessment of task performance.

- **Identification of Exceptions.** An exceptions report should be included with the task performance scores and risk scores to explain and highlight deviations in the scores. It provides a means to show details on areas of concern. A sample format for an exceptions report is illustrated as table 2-22.

**Table 2-22. Task Performance Exceptions**

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
ID No.	Task	Condition	Attribute	Measure	Exception
T1	2	1	2	3	Issue
T2	2	2	2	1	Issue
T3	3	1	3	2	Issue
T4	3	2	3	2	Issue

## Mission Effectiveness Assessment

### Overview

Mission effectiveness assessment may be quantitative or qualitative. A quantitative assessment makes use of measures developed to assess the achievement of mission desired effects. A qualitative assessment recognizes the difficulty in evaluating MLMs and makes use of sub-COIs focused on mission desired effects to assess SUT impact on mission effectiveness. Ideally, it is much more desirable to conduct a quantitative assessment of mission effects, but often it is just not feasible.

### Process

- **Quantitative Process.** The quantitative process for assessing mission effectiveness follows the same process as that used for assessing task performance. The only differences are:
  - Missions are used instead of tasks.
  - Attributes are based on mission desired effects.
  - Weighting of desired effects ( $W_e$ ) is added to the scoring and risk process.

The scoring calculation for mission **M** would then be:

The risk calculation for mission **M** would be:

Similar tables as those in 2-18 through 2-21 would be constructed for the mission assessment process using the appropriate mission level data.

- **Qualitative Process.** Given that mission measures data is not available or is limited, the risk of assessing mission effectiveness quantitatively may be too large. A qualitative process may be the preferred option. As discussed in the SOP, a COI is developed for each mission. Sub-COIs are developed for mission desired effects. These sub-COIs are put in terms of assessing the SUT impact on the desired effect.

A mission level assessment should support the warfighter in determining if the SUT satisfies the gap that was originally identified. The warfighter will typically say he wants to do something better or faster. The mission assessment should indicate how the SUT supports doing the mission better or faster.

A qualitative assessment means that there is little to no measurable data and that the assessment is mostly based on subjective relationships and insights. This is where the relationship mapping in table 2-2 between mission attributes and task attributes becomes helpful. Deficiencies in task performance attributes will tend to impact those related mission level attributes. Higher weighted attributes will have a greater impact than lesser weighted attributes.

A quantitative assessment may also include surveys of SMEs and field experienced warfighters. Subjective insights may be determined based on this technique and included in the mission effectiveness assessment.

Table 2-23 illustrates a sample format for presenting a qualitative assessment of a single mission. This assessment should be done for each mission and be included in any reports on the SUT.

**Table 2-23. Sample Format for a Qualitative Assessment of Mission Effects**

<b>Mission COI:</b> Can the SUT support mission One?
<b>Sub-COI:</b> Assess SUT ability to support mission desired effect 1
Attribute 1
Insights
Attribute 2
Insights
<b>Sub-COI:</b> Assess SUT ability to support mission desired effect 2
Attribute 1
Insights
Attribute 3
Insights
<b>Sub-COI:</b> Assess SUT ability to support mission desired effect 3
Attribute 2
Insights
Attribute 3
Insights

## STEP 3: REPORTING STANDARD

Step 3 establishes a standard for reporting the assessment of SUT effectiveness and suitability in terms of system/SoS functionality, task performance, and mission effectiveness. The purpose of reporting is to provide a logical presentation of findings and conclusions that will enable the test and evaluation authorities to justify the results and to support the determination on future acquisition of the SUT. Once at the reporting phase of the process, the test team should have completed the right side of the “V” diagram in figure 1-1.

### **System/SoS Assessment Reporting**

Due to the potential for a large number of system/SoS attributes, the system/SoS assessment reporting needs to be a summary that highlights key deficiencies in the SUT functionality. Table 2-24 illustrates one possible method to present the information on the system. There are three major sections to this format.

- The first section provides specific information on each effectiveness and suitability KPP. Color-coded scores are provided for KPP measurement and risk. Exceptions are identified by number and shown in section three.
- The second section is devoted to summarizing the effectiveness and suitability KSAs and other attributes. Instead of providing actual scores for each, a count is provided on “how many passed” and “how many failed” to meet their criteria. Instead of individual risk scores, the risk scores for each category of attributes can be reported as an average. Exceptions are identified by number and shown in section 3.
- Section 3 provides the exceptions report. Exceptions are based on effectiveness and suitability attributes. Issue statements in the exceptions report can be either positive or negative.

### **Task Performance Reporting**

Reporting on task performance provides an operational context for evaluating warfighter requirements. Table 2-25 illustrates one possible method to present the information. There are three major sections to this format.

- The first section provides summary information on the tasks for each mission the SUT is designed to support. Color-coded scores are provided for both task measurement and risk. Exceptions are identified by number and shown in section 3.
- The second section is devoted to summarizing the tasks performed by the SUT and those tasks that the SUT supports (provides input to). Instead of providing actual scores for each, a count is provided on “how many passed” and “how many failed” to meet their criteria. Instead of individual risk scores, the risk scores for each category of tasks are reported as an average. Exceptions are identified by number and shown in section 3.
- Section 3 provides the exceptions report. Exceptions are based on issues related to task performance. Issue statements in the exceptions report can be either positive or negative, reflecting strengths and weaknesses in task performance and/or the ability to evaluate the task.

Table 2-24. Sample System/SoS Assessment Report

SECTION 1					
Attribute Category	Score		Risk Score	Exceptions	
Effectiveness KPPs					
KPP 1	0.915		0.272		
KPP 2	0.4		0.385	E1	
Suitability KPPs					
KPP 3	0.8		0.2	S1	
KPP 4	0.85		0.25		
SECTION 2					
Attribute Category	Count	Ct Passed	Ct Failed	Avg Risk Score	Exceptions
Effectiveness Attributes					
KSAs	2	2	0	0.216	E2, E3
OAs	2	2	0	0.06	
Suitability Attributes					
KSAs	3	3	0	0.2	S2
OAs	4	4	0	0.15	S3
SECTION 3					
Exceptions - Effectiveness					
ID No.	Attribute		Type	Exception	
E1	1		KPP	Issue statement	
E2	3		KSA	Issue statement	
E3	4		KSA	Issue statement	
Exceptions - Suitability					
ID No.	Attribute		Type	Exception	
S1	7		KPP	Issue statement	
S2	8		KSA	Issue statement	
S3	9		OA	Issue statement	

**Table 2-25. Sample Task Performance Assessment Report**

SECTION 1					
Mission	Score			Risk Score	Exceptions
Mission 1 Tasks	0.758			0.054	T1 – T4
Mission 2 Tasks	0.8			0.1	T5 – T7
Mission 3 Tasks	0.9			0.2	T8
SECTION 2					
Tasks	Count	Count Passed	Count Failed	Avg Risk Score	Exceptions
Mission 1					
Tasks performed by SUT	3	2	1	0.053	T1, T2
Tasks supported by SUT	2	2	0	0.055	T3, T4
Mission 2					
Tasks performed by SUT	4	4	0	0.13	T5, T6
Tasks supported by SUT	3	3	0	0.08	T7
Mission 3					
Tasks performed by SUT	5	5	0	0.22	T8
Tasks supported by SUT	1	1	0	0.18	T8
SECTION 3					
Mission 1					
ID No.	Task	Attribute		Exception	
T1	2	2		Issue statement	
T2	2	2		Issue statement	
T3	3	3		Issue statement	
T4	3	3		Issue statement	
Mission 2					
ID No.	Attribute	Attribute		Exception	
T5	6	7		Issue statement	
T6	7	7		Issue statement	
T7	8	9		Issue statement	
Mission 3					
ID No.	Attribute	Attribute		Exception	
T8	12	10		Issue statement	

### Mission Effectiveness Reporting

Mission effectiveness reporting will depend on the type of assessment at the mission level. If a quantitative assessment is conducted in which MLMs are evaluated, then the mission effectiveness report may look similar to table 2-17 for task performance. If the assessment is qualitative, based on COIs and sub-COIs, then the mission effectiveness report may look like table 2-23. Exception reporting may be included to explain deficiencies in the mission.

## Root Cause Analysis

Since the measures framework evaluates mission, tasks, and system attributes separately and with their own measures, it is not always evident how the system impacts task performance and mission effectiveness. The assessment process requires determining the cause or causes for shortcomings in task performance and mission effectiveness by linking the assessments of system, tasks, and mission. Mission deficiencies may be based on task performance deficiencies, which may be based on system/SoS deficiencies. The linkages occur through attributes. System-task relationship mapping and task-mission relationship mapping (see table 2-2) provide the linkages needed to analyze cause and effect.

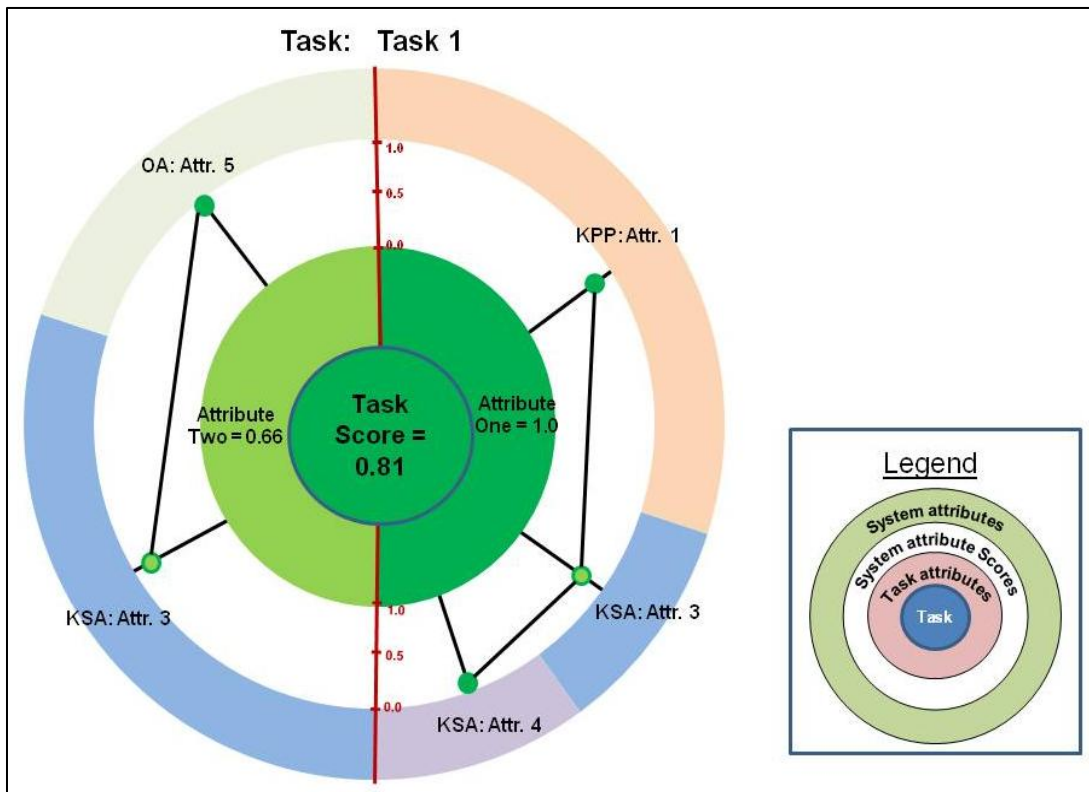
Table 2-26 illustrates the linkages that will inform the root cause analysis. This table is based on previously constructed tables (2-2, 2-5, 2-13, and 2-18) to show potential cause and effect relationships. The color-coded scores for the tasks, task attributes, and system attributes enable the observer to quickly see where the deficiencies are and possible causes. For example, it may be seen that task 2 is deficient due to task attribute 3 which may be due to system attribute 2 – KPP.

**Table 2-26. Mission – Task - System Linkages**

Mission	Desired Effects (Sub-COIs)	Tasks	Task Score	Task Attributes	Task Attribute Score	System Attributes	System Attribute Score
Mission 1	(1) Sub-COI 1 (2) Sub-COI 2 (3) Sub-COI 3	1	0.81	1	1.0	KPP: Attribute 1	0.915
						KSA: Attribute 3	0.80
						KSA: Attribute 4	1.0
		2	0.44	2	0.66	KSA: Attribute 3	0.80
						OA: Attribute 5	1.0
						KPP: Attribute 1	0.915
		3	0.75	3	0.0	KSA: Attribute 3	0.80
						OA: Attribute 5	1.0
						KPP: Attribute 2	0.40
		4	1.0	1	1.0	KSA: Attribute 4	1.0
						KSA: Attribute 4	1.0
						OA: Attribute 6	1.0
		5	0.81	3	0.0	KPP: Attribute 1	0.915
						KSA: Attribute 3	0.80
						KPP: Attribute 2	0.40
		5	0.81	3	0.0	KSA: Attribute 4	1.0
						KSA: Attribute 4	1.0
						OA: Attribute 6	1.0



Figure 2-3 provides an alternative means to illustrate system task causal relationships. This may be used in addition to table 2-26 or as a replacement. However, this figure represents only the relationships for task 1. A separate figure would be needed for each task that is evaluated. The center circle shows the overall task 1 score and is color coded. The inner-circle then shows the three task attributes and their scores. Each one is also color coded. The size of the wedges for each task attribute represents the weighted priority for each attribute. The outer circle represents the system attributes that are linked to each task attribute. In this illustration, their size varies based on their relative weight to the task attribute. However, since this is not already determined in the assessment process, the weighting may not be available and, therefore, equally weighted and sized in the figure. The white area between the outer and inner rings provides a scaled area to show the score for each system attribute. Several causal relationships may be surmised from this diagram. First, that task attribute 1 appeared to have little to no impact by system KPPs and KSAs scoring less than 1. Second, that task attribute 2 appeared to have had some impact from KSA attribute 3.



**Figure 2-4. System – Task Causality Diagram**

Table 2-27 provides a summary of key issues determined by a root cause analysis. This may be used to summarize findings based on causal relationships.

**Table 2-27. Root Cause Analysis**

System/SoS Functions			Task Performance				Mission Effects	
Attribute	Issue		Attribute	Task	Issue		Attribute	Desired Effects
KSA Attr. 3	SUT Issue 1	<b>Degraded</b> →	Task Attr. 2	Task 1	Task Issue 1	<b>Degraded</b> →	Mission Attr. 1	DE 1
KPP Attr. 2	SUT Issue 2	<b>Degraded</b> →	Task Attr. 3	Task 2	Task Issue 2	<b>Degraded</b> →	Mission Attr. 2	DE 2

## CHAPTER 3

### ASSESSMENT PROCESS EXAMPLE

#### INTRODUCTION

This chapter provides an example of the assessment process outlined in chapter two. The example will use the PLB described in the Measures Development SOP as the SUT supporting the JPR mission thread. The matrixes from the SOP are included in annex C of this guidebook for reference. The weightings and data values found in this example are arbitrary and not based on actual data. Therefore, the information and results provided in this chapter are for illustration purposes only.

## STEP 1: INITIATING THE ASSESSMENT PROCESS

This step will validate the work from the Measures Development SOP and fill in the gaps as needed.

### Relationship Mapping

#### Overview

A verification of relationship mapping is shown in table 3-1. As the table illustrates, two relationships are not mapped, and, therefore, mapping is required before proceeding to the next step.

**Table 3-1. Required Relationships for Mission-Based Assessment**

Level	Relationship	Completed	Location
Mission	Conditions to Mission	Missing	
Mission	Desired effects to Attributes	X	Figure C-2
Mission	Attributes to Measures	X	Figure C-3
Task	Tasks to attributes	X	Figure C-6
Task	Attributes to Measures	X	Figure C-7
System/SoS	SUT to attributes	X	Figure C-8
System/SoS	Attributes to measures	X	Figure C-10
System - Task	System/SoS Attributes to Task Attributes	X	Figure C-9
Task - Mission	Task Attributes to Mission Attributes	Missing	

## Process

The process for mapping relationships does not have to be complex. The simple process is to establish a one-to-one relationship. A many-to-one relationship mapping becomes more complex and may require detailed input from SMEs.

- **Conditions to Mission.** Since only one mission (JPR) exists in this example, the relationship mapping is simply an identification of condition sets for which the mission will be performed. Two simple environmental condition sets are chosen for the JPR mission as shown in table 3-2.

**Table 3-2. Condition Sets for JPR Mission**

Mission	Condition Set Descriptions	
JPR	<u>Condition Set 1:</u> Harsh tropical environment with hostile forces scattered throughout the area	<u>Condition Set 2:</u> Mountainous cold climate with limited line of sight connectivity

- **Task Attributes to Mission Attributes.** Mapping task attributes to mission attributes requires identifying those attributes at each level. Referring to figures C-2 and C-6, the mission and task attributes can be listed in table 3-3. An understanding of each attribute description will assist in determining the relationships. Common attributes at the mission and task levels will typically be related (for example, timeliness at task level will normally map to timeliness at mission level). Understanding mission attributes and their associated desired effects will also help in identifying relationships. Note that in table 3-3 there are no task attributes mapped to the mission attribute of “Readiness.” That is because the task supported by the SUT is “Locate IP” (isolated personnel). This task is part of the execution phase, whereas the “Readiness” attribute is related to a desired effect that is focused on the preparation phase of the JPR mission thread.

**Table 3-3. Task Attributes to Mission Attributes Relationship**

	Mission Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness
Task - Attributes						
Accuracy			X			
Timeliness			X			
Information Reliability				X	X	
Completeness		X			X	

# Prioritization

## Overview

Prioritization will help to characterize real world warfighter requirements so that the assessment process can duplicate relative importance. Chapter 2 of this guidebook discussed the concepts and shortcuts for conducting a prioritization. Using those guidelines, previous developed tables will need to quantify priorities. Table 3-4 lists the relationships that require priorities to support the assessment process.

**Table 3-4. Required Relationships Prioritizations**

Element	Relationship Description	Supporting Tables
Mission	Prioritize missions conducted by the SUT	
Conditions	Prioritize conditions for each mission	Figure 3-2
*Mission Desired Effects	Prioritize desired effects for each mission	Figure C-1
*Mission Attributes	Prioritize mission attributes for each mission desired effect	Figure C-2
*Mission Measures	Prioritize mission measures for each mission attributes	Figure C-3
Tasks/Sub-tasks	Prioritize tasks and sub-tasks	Figure C-5
Task Attributes	Prioritize task attributes for each task	Figure C-6
Task Measures	Prioritize task measures for each task attribute	Figure C-7
SUT/SoS Attributes	Prioritize SUT/SoS Attributes	Figure C-8
SUT/SoS Measures	Prioritize SUT/SoS measures for each attribute	Figure C-10

\*NOTE: Not required if mission assessed through COIs and sub-COIs (mission measures not evaluated).

## Process

- Mission Level Prioritizations.** In assessing mission impacts, of the SUT, condition sets, mission desired effects, mission attributes, and mission measures may need to be prioritized. If the mission level assessment will be based on COIs and not on actual MLMs, then the latter three elements are not required to be prioritized. Each element will be address separately, but the information for a mission can be summarized into a single table similar to that shown in table 2-4.
  - Mission Prioritization.** In this example, only one mission (JPR) is used. Therefore, the weighting for the mission is 100% and will not impact the assessment process.
  - Condition Prioritization.** Table 3-2 established that there will be two condition sets in which the SUT will be evaluated. It is believed that each condition set is likely to occur, and, therefore, each condition set will be weighted 50% as shown in table 3-5.

**Table 3-5. Condition Sets Weightings for JPR Mission**

Mission	Condition Set Descriptions	Weighting
JPR	<u>Condition Set 1:</u> Harsh tropical environment with hostile forces scattered throughout the area	50%
	<u>Condition Set 2:</u> Mountainous cold climate with limited line of sight connectivity	50%

- **Mission Desired Effects Prioritization.** Figure C-1 will be used as a basis for prioritizing JPR desired effects. Since these are mapping to mission objectives, a simple scoring model can be used to determine weights. It is assumed the objectives are in listed in order of priority. Assigning weights to them with the highest number having the most weight, the mapped attributes can be weighted. Table 3-6 shows the resulting calculations and weights for the JPR desired effects.

**Table 3-6. JPR Desired Effects Weight Calculations**

Weight	Objectives	Desired Effects	Return Isolated Personnel to Duty	Sustain Morale	Increase Operational Performance	Deny Adversary Ability to Exploit the IP
3	Protect the force		3	3		
2	Enable military missions		2	2	2	
1	Defeat adversary attempts to exploit a known asymmetric vulnerability		1			1
Column Sum (CS)			6	5	2	1
Total Sum (TS)			14	14	14	14
Weight (CS/TS)			0.43	0.36	0.14	0.07

- **Mission Attributes Prioritization.** Attributes are weighted for each desired effect based on the relative influence each attribute has on that desired effect. The attribute weights for each desired effect must total 100%. In most cases, only one or two attributes may exist for a desired effect. A single attribute will be weighted 100% while two attributes may be weighted equally 50% or vary based on priority. The weights for the JPR mission example are shown in table 3-7.

**Table 3-7. JPR Mission Attribute Weights**

Desired Effect	Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness	Totals
Return isolated personnel to duty		100%					100%
Sustain morale			100%				100%
Increase operational performance				50%	50%		100%
Deny adversary opportunity to exploit the IP					80%	20%	100%

- **Mission Measures Prioritization.** Mission measures are weighted for each attribute they measure, and weight is based on a relative value to the attribute. The measure weights for each attribute must total 100%. In many cases, only one measure will exist for a single attribute. More than one measure will need to be weighted. The weights for the JPR mission example are shown in table 3-8.

**Table 3-8. JPR Mission Measure Weights**

Scale	Mission Measures	Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness
Percent	Of JPR missions where IP personnel was available to be cleared for duty		50%				
Percent	Of JPR missions where IP personnel was cleared for duty		50%				
Time	For IP to respond to changes in threat and environmental conditions that required the IP to evade, resist, or escape			50%			
Percent	Of JPR missions where morale was a factor in IP inability to survive, evade, resist, or escape			50%			
Percent	Of JPR missions where no unplanned redundant activities occurred				25%		
Percent	Of JPR missions where continuous horizontal coordination existed across operational nodes				25%		
Percent	Of JPR missions where continuous vertical coordination existed across operational nodes				25%		
Percent	Of JPR mission executions where planning and preparation led to successful coordination across operational nodes				25%		
Percent	Of JPR missions where correct decisions were made by operational nodes based on situational awareness					40%	
Percent	Of JPR missions where IP acted correctly based on situational awareness					40%	
Percent	Of JPR missions where operational nodes acted correctly based on situational awareness					20%	
Percent	Of JPR missions where inadequate training led to mission execution deficiencies						50%
Percent	Of JPR missions where inadequate systems, supplies, and resources led to mission execution deficiencies						50%
Totals			100%	100%	100%	100%	100%

- **Task Level Prioritizations.** At the task level, tasks and sub-tasks will need to be prioritized, as will their attributes and measures. Each element will be address separately, but can be summarized into a single table similar to that shown in figure 2-10.
  - **Task and Sub-Task Prioritization.** Tasks and sub-tasks will usually be of equal importance in a mission thread as each is needed to perform the mission. However, when looking at a segment of the mission thread, they may not be equally weighted in the assessment. Table 2-6 provided a weighting schema to use that is based on how the SUT relates to the task (sub-task). This example focuses only on the “Locate” task in the JPR mission thread, as this is where the PLB (the SUT) is utilized. Table 3-9 shows the relevant tasks and sub-task along with the weighting calculations based on the table 2-6 schema. In this example, the SUT is not the primary performer in each of the “Locate” sub-tasks, but does provide input as a supporting system to the first three sub-tasks. Therefore, sub-tasks 4.1 to 4.3 are given a score of 1.0 and have a task weight of 33.3%. Chapter 1 discussed when it is feasible to evaluate a set of sub-tasks as a single task. In this example, if all four sub-tasks for the “Locate” task were supported by the SUT, then it would have been feasible to assess the SUT impact on task performance by evaluating measures at the “Locate” task level and not at the sub-task level. This would have

simplified the assessment to only evaluate one task with a smaller set of measures. However, that is not the case.

**Table 3-9. “Locate” Tasks Weights**

Col 1	Col 2	Col 3	Col 4	Col 5
Task No.	Tasks Sub-Tasks	Description	Schema Weights*	Task Weight
4.0	Locate			
4.1	Execute Search Plan	Determine the location and status of the isolated personnel (precisely find; fast response)	1.0	33.3%
4.2	Verify/Fuse Location	Verify and fuse isolated personnel's location information to provide accurate and reliable coordinates for refining recovery plans. Goal is for latest, most reliable location information.	1.0	33.3%
4.3	Authenticate IP	Authenticate isolated personnel using Isolated Personnel Report (ISOPREP) data and other methods	1.0	33.3%
4.4	Share Location	Use available information to refine isolated personnel's location with reliable and accurate information	0.0	0%
Column Total			4.0	

\*Schema weights based on categories established in table 2-6.

- o **Task Attributes Prioritization.** Attributes are weighted for each task based on the relative influence each attribute has on that task. The attribute weights for each task must total 100%. In most cases, only one or two attributes may exist for a task. A single attribute will be weighted 100% while two attributes may be weighted equally 50% or vary based on priority. The weights for the JPR “Locate” task relevant to the SUT example are shown in table 3-10.

**Table 3-10. Task Attribute Weights**

Tasks	Attributes	Accuracy	Timeliness	Information Reliability	Completeness	Totals
Execute Search Plan		33%	33%	34%		100%
Verify/Fuse Location		33%	33%	34%		100%
Authenticate IP		50%	50%			100%
Share Location		33%	33%		34%	100%

- **Task Measures Prioritization.** Task measures are weighted for each attribute they measure with the weighting based on a relative value to the attribute. The measure weights for each attribute must total 100%. In many cases, only one measure will exist for a single attribute. More than one measure will need to be weighted. The weights for the JPR “Locate” task relevant to the SUT example are shown in table 3-11.



**Table 3-11. JPR “Locate” Task Measure Weights**

Task	Scale	Task Measures	Attributes	Accuracy	Timeliness	Information Reliability	Completeness
Execute Search Plan	Percent	Of instances where search was executed according to plan before the IP was found		100%			
	Time	To commence search from time tasked			100%		
	Count	Of search passes over IP position before IP located				100%	
	Totals:			100%	100%	100%	
Verify/Fuse Location	Percent	Of instances where search data was accurately verified and fused with prior data resulting in correct determination		100%			
	Time	For search data to be verified and fused with prior data which result in a correct determination			100%		
	Percent	Of instances where reliability of the data was maintained or improved from the verify and fuse process				100%	
	Totals:			100%	100%	100%	
Authenticate IP	Percent	Of instances where the authentication of the IP to the location was accurate with real truth information		100%			
	Time	To authenticate IP from time first located			100%		
	Totals:			100%	100%		
Share Location	Percent	Of shared information exchanges where information received was accurate with what was sent		100%			
	Percent	Of shared information exchanges where information received was complete when compared to what was sent				100%	
	Time	To execute the share location sub-task where information exchange was accurate and complete		100%			
	Totals:			100%	100%		100%

- **SUT/SoS Functionality Prioritization.** SUT/SoS functionality is assessed through system/SoS attributes and system/SoS level measures. Attributes are already prioritized by attribute type (KPP, KSA, or other attributes). System/SoS measures will need to be prioritized based on a relative value to the attribute. Conditions are also important when evaluating system/SoS attributes. Each element will be addressed separately, but can be summarized into a single table similar to that shown in table 2-7.
  - **Condition Sets for System/SoS Attributes.** System/SoS attributes may function differently based on environmental and threat conditions. Condition sets need to be

included in determining system/SoS attribute and measures priorities. Recall from chapter 2 that since a SUT may function under several mission and condition sets, the weighting of each condition set must be determined with the following formula:

Since only one mission exists in this example,  $W_M = 1$  and, therefore, the calculation is simple as shown in table 3-12. A second non-existent mission is included in the table just to illustrate how to calculate the condition set weights.

**Table 3-12. Condition Sets Weightings for SUT Attributes**

Column 1	Column 2	Column 3	Column 4
Missions		Condition Sets	
		Condition Set 1: Harsh tropical environment with hostile forces scattered throughout the area	Condition Set 2: Mountainous cold climate with limited line of sight connectivity
<b>Weight</b>	<b>Title</b>		
100%	JPR	50%	50%
0	XX	50%	50%
<b>Totals</b>		<b><math>(1 \cdot .5) + (0 \cdot .5) = 0.5</math></b>	<b><math>(1 \cdot .5) + (0 \cdot .5) = 0.5</math></b>

- o **System/SoS Attributes Prioritization.** The system/SoS attribute priorities are based on attribute type following the schema illustrated in table 2-8 and as shown in table 3-13. Notice that this includes both effectiveness and suitability attributes as 1.0 in the weighting schema. It is possible to consider them separately.

**Table 3-13. System/SoS Attribute Weighting Schema**

Column 1	Column 2	Column 3	Column 4	Column 5
Attribute Category	Count of Attributes in Category	Schema Weight	Total Weight (Col 2 * Col 3)	Calculated Weight for Each Attribute (Col 3/Sum)
KPP	2	3.0	6	0.1875
KSA	3	2.0	6	0.125
Other Attribute	4	1.0	4	0.0625
<b>Sum</b>			<b>16</b>	

- o **System/SoS Measures Prioritization.** System/SoS measures are weighted for each attribute they measure, and weight is based on a relative value to the attribute. The measure weights for each attribute must total 100%. In some cases, only one measure will exist for a single attribute. If there is more than one measure for an attribute, then those measures will need to be weighted. The weights for the PLB SUT attribute measures are shown in table 3-14. Notice that there are sub-attributes defined for some

of the attributes. Column 3 shows the weighting of each sub-attribute (column 4) to the main attribute (column 2). Column 5 is the weighting of the measure (column 7) to the sub-attribute (column 4).

**Table 3-14. System/SoS Measures Weighting Schema**

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
Type	Attribute	Weight	Sub-Attribute	Weight	Scale	Measures
KPP	Operational Capability	100%	Range	100%	NM	Max range for clear continuous signal
KPP	Net-Ready	100%	Interoperability	100%	Percent	SAR systems interoperable with
KSA	Protection	50%	Transmitted Data Accuracy	50%	Percent	Data transmissions that are complete
				50%	Percent	Data transmissions that are complete and accurate
		50%	Access and Control	100%	Y/N	Single handed controllable operations
KSA	Sustainment	50%	Reliability	100%	Percent	Probability operable for 24 hr period
		50%	Ownership Cost	100%	\$\$\$	Annual maintenance cost
KSA	Interoperability	100%	Transmission Output	100%	Percent	Continuous transmission power output
OA	N/A	100%	Shock Resistant	100%	Percent	Operable after ejected from aircraft seat
OA	N/A	100%	Speed of Initial Report	100%	Seconds	Time between activation and initial beacon broadcast
OA	N/A	100%	Water Resistant	100%	Meters	Max depth maintains watertight
OA	N/A	100%	Battery Life	100%	Years	Max battery shelf life

## STEP 2: CONDUCT THE ASSESSMENT

This example will continue to use the Threshold Model in annex D for single measure scoring and various other models and color codes illustrated in chapter 2. The models that will be used in this example are listed in table 3-15. Taking a bottom-up approach to assessment, the system/SoS attributes are assessed first, then task performance, and finally mission effects assessment.

**Table 3-15. Assessment Models**

Model	Reference
Threshold Model for Single Measure Scoring	Annex D
Aggregate Measure Scoring Model	Table 2-9
Risk Matrix	Figure 2-2
Risk Color Codes	Table 2-10
Likelihood Levels	Table 2-11
Impact Levels	Table 2-12

### System/SoS Assessment

#### Overview

The purpose of the system/SoS evaluation is to determine the functional capabilities of the system/SoS when employed in a realistic operational environment. The system/SoS effectiveness and suitability attributes will be grouped together in this example.

#### Process

This example will follow the process illustrated in chapter 2 to include evaluating data, determining scores for each measure and the measure reliability (risk), consolidating the measures and attributes, and identifying exceptions.

- **Evaluation of Data.** The evaluation is assumed to have occurred based on standard evaluation processes and techniques. Evaluation results shall be compared to measure threshold and objective values.
- **Scoring Measures.** Each measure for system/SoS effectiveness and suitability attributes shall be scored based on the Threshold Model in annex D. The results are summarized in table 3-16. Column 12 provides an aggregated score for each KPP, KSA, and other attribute.

**Table 3-16. PLB System/SoS Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11	Col 12
Type	Attribute	Sub-Attribute	Measure Weight (W <sub>m</sub> )	Measure	Threshold value (TV)	Objective value (OV)	Condition Set	Condition Weight (W <sub>c</sub> )	Observed value	Score value (S <sub>cm</sub> )	Total Score (S <sub>a</sub> )
KPP	Operational Capability	Range	100%	Max range for clear continuous signal	50 NM	100 NM	One	100%	75NM	1	1.0
KPP	Net-Ready	Interoperable	100%	Pct SAR systems interoperable with	100% US	100% US and NATO	One	100%	95%	0	0.0
KSA	Protection	Transmitted Data Accuracy	25%	Data transmissions that are complete	99%	Same	One	100%	99%	1	0.75
			25%	Data transmissions that are complete and accurate	99%	Same	One	100%	95%	0	
		Access and Control	50%	Is the PLB single hand controllable operations	Yes	Same	One	100%	Yes	1	
KSA	Sustainment	Reliability	50%	Pct probability operable for 24 hr period	95%	99%	One	50%	96%	1	0.75
					95%	99%	Two	50%	90%	0	
		Ownership cost	50%	Annual maintenance cost in dollars	\$50 annual	\$25 annual	One	100%	\$45	1	
KSA	Interoperability	N/A	100%	Pct continuous transmission power output	20W	25W	One	50%	25W	1	0.5
					20W	25W	Two	50%	24W	0	
OA	Shock Resistant	N/A	100%	Pct operable after ejected from A/C seat	99%	100%	One	100%	96%	0	0.0
OA	Speed of Initial Report	N/A	100%	Time between activation and initial beacon broadcast	5 sec	2 sec	One	100%	2 sec	1	1.0
OA	Water Resistant	N/A	100%	Max depth maintains watertight	5 M	10 M	One	100%	7 M	1	1.0
OA	Battery Life	N/A	100%	Max battery shelf life	5 Yr	7 Yr	One	100%	5 Yr	1	1.0

- **Measuring Risk.** Each measure is assessed for risk of error in determining if it met the threshold value. Table 3-17 provides an example for the PLB SUT. Column 13 provides the impact values from table 2-12. Column 14 provides the likelihood or error based on the data analysis and table 2-11. Column 15 provides the calculated risk for each attribute, and column 16 provides the risk score (Col 15 ÷ 25).

**Table 3-17. PLB System/SoS Risk Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 8	Col 9	Col 12	Col 13	Col 14	Col 15	Col 16
Type	Attribute	Sub-Attribute	Measure Weight (W <sub>m</sub> )	Measure	Condition Set	Condition Weight (W <sub>c</sub> )	Total Score (S <sub>a</sub> )	Impact (I <sub>a</sub> )	Likelihood of error (L <sub>cm</sub> )	Calc. Risk	Risk Score (R <sub>a</sub> )
KPP	Operational Capability	Range	100%	Max range for clear continuous signal	One	100%	1.0	5	3	15	0.60
KPP	Net-Ready	Interoperable	100%	Pct SAR systems interoperable with	One	100%	0.0	5	1	5	0.20
KSA	Protection	Transmitted Data Accuracy	25%	Data transmissions that are complete	One	100%	0.75	3	2	4.5	0.18
			25%	Data transmissions that are complete and accurate	One	100%			2		
		Access and Control	50%	Is the PLB single hand controllable operations	One	100%			1		
KSA	Sustainment	Reliability	50%	Pct probability operable for 24 hr period	One	50%	0.75	3	2	11.25	0.45
					Two	50%			3		
		Ownership cost	50%	Annual maintenance cost in dollars	One	100%			5		
KSA	Interoperability	N/A	100%	Pct continuous transmission power output	One	50%	0.5	3	1	3	0.12
					Two	50%			1		
OA	Shock Resistant	N/A	100%	Pct operable after ejected from A/C seat	One	100%	0.0	1	4	4	0.16
OA	Speed of Initial Report	N/A	100%	Time between activation and initial beacon broadcast	One	100%	1.0	1	1	1	0.04
OA	Water Resistant	N/A	100%	Max depth maintains watertight	One	100%	1.0	1	2	2	0.08
OA	Battery Life	N/A	100%	Max battery shelf life	One	100%	1.0	1	5	5	0.20

- **Aggregating Scores.** Aggregated scores for system/SoS attributes may be determined for both attribute measures and attribute risk.
  - **Aggregate Attribute Measures.** Tables 3-13 and 3-16 provide the information needed to aggregate attribute measures into one score. Using values from column 5 of table 3-13 as weights and scores from column 12 of table 3-16, a calculated aggregate score is shown in table 3-18. Based on the color codes for scores in table 2-9, the aggregated score for system/SoS attributes is light green. Note that even though the aggregate score is light green, failure of a KPP may dictate that the SUT fails the T&E.

**Table 3-18. Aggregate System/SoS Scoring Table**

Table 3-13	Table 3-16			New Column
Column 5	Column 1	Column 2	Column 12	Aggregate Score ( $S_{SUT}$ )
Calc. Attribute Weight ( $W_a$ )	Type	Attribute	Score ( $S_a$ )	
0.1875	KPP	Operational Capability	1.0	0.625
0.1875	KPP	Net-Ready	0.0	
0.125	KSA	Protection	0.75	
0.125	KSA	Sustainment	0.75	
0.125	KSA	Interoperability	0.5	
0.0625	OA	Shock Resistant	0.0	
0.0625	OA	Speed of Initial Report	1.0	
0.0625	OA	Water Resistant	1.0	
0.0625	OA	Battery Life	1.0	

- **Aggregate Attribute Risk.** Tables 3-13 and 3-17 provide the information needed to aggregate attribute risk into one score. Using values from column 5 of table 3-13 as weights and risk scores from column 16 of table 3-17, a calculated aggregate score is shown in table 3-19. Based on the color codes for scores in table 2-10, the aggregated risk score for the system/SoS attributes is green.

**Table 3-19. Aggregate System/SoS Risk Scoring Table**

Table 3-13	Table 3-17			New Column
Column 5	Column 1	Column 2	Column 16	Aggregate Risk Score ( $R_{SUT}$ )
Calc. Attribute Weight ( $W_a$ )	Type	Attribute	Risk Score ( $R_a$ )	
0.1875	KPP	Operational Capability	0.60	0.274
0.1875	KPP	Net-Ready	0.20	
0.125	KSA	Protection	0.18	
0.125	KSA	Sustainment	0.45	
0.125	KSA	Interoperability	0.12	
0.0625	OA	Shock Resistant	0.16	
0.0625	OA	Speed of Initial Report	0.04	
0.0625	OA	Water Resistant	0.08	
0.0625	OA	Battery Life	0.20	

- **Identification of Exceptions.** The exceptions report shown in table 3-20 helps to explain deficiencies with the system/SoS attribute scores and risk.

**Table 3-20. System/SoS Attribute Exceptions**

Column 1	Column 2	Column 3	Column 4
<b>Attribute Scores</b>			
ID No.	Attribute	Type	Exception
E1	Net-Ready	KPP	Not interoperable with P-3
E2	Protection	KSA	Some data packets dropped over time
E3	Sustainment	KSA	Reliability affected in cold weather
E4	Interoperability	KSA	Power output affected in cold weather
E5	Shock Resistant	OA	Failure after several shock tests
<b>Attribute Risk Scores</b>			
ID No.	Attribute	Type	Exception
E6	Operational Capability	KPP	Limited data not statistically significant
E7	Sustainment	KSA	Maintenance costs estimated over time
E8	Battery Life	OA	Unable to fully test

## Task Performance Assessment

### Overview

The purpose of the task performance assessment is to determine whether the SUT and its operators support the SoS in providing the necessary capabilities to perform tasks.

### Process

This process includes evaluating data, determining scores for each measure and the measure reliability (risk), aggregating the measure and risk scores, and identifying exceptions.

- **Evaluation of Data.** The evaluation of data is assumed to have already occurred based on standard evaluation processes, tools, and techniques. Evaluation results shall be compared to measure threshold and objective values.
- **Scoring Measures.** Each measure for task performance shall be scored based on the Threshold Model in annex D. The results are summarized in table 3-21. Column 12 provides an aggregated score for each sub-task. The results are color coded based on table 2-9.



**Table 3-21. Task Performance Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11	Col 12
Task	Condition Weight (W <sub>c</sub> )	Condition Set	Attribute Weight (W <sub>a</sub> )	Attribute	Measure Weight (W <sub>m</sub> )	Measure	Threshold value (TV)	Objective value (OV)	Observed value (ObV)	Score value (S <sub>cam</sub> )	Total Score (S <sub>t</sub> )
Execute Search Plan	50%	1	33%	Accuracy	100%	Percent of instances where search was executed according to plan before the IP was found	90%	95%	92%	1	0.66
			33%	Timeliness	100%	Time to commence search from time tasked	10 min	5 min	9 min	1	
			34%	Information Reliability	100%	Count of search passes over IP position before IP located	2	1	3.5	0	
	50%	2	33%	Accuracy	100%	Percent of instances where search was executed according to plan before the IP was found	90%	95%	92%	1	
			33%	Timeliness	100%	Time to commence search from time tasked	10 min	5 min	9 min	1	
			34%	Information Reliability	100%	Count of search passes over IP position before IP located	2	1	2.5	0	
Verify/Fuse Location	50%	1	33%	Accuracy	100%	Percent of instances where search data was accurately verified and fused with prior data resulting in correct determination	90%	100%	95%	1	0.84
			33%	Timeliness	100%	Time for search data to be verified and fused with prior data which result in a correct determination	10 min	5 min	10 min	1	
			34%	Information Reliability	100%	Percent of instances where reliability of the data was maintained or improved from the verify and fuse process	90%	95%	95%	1	
	50%	2	33%	Accuracy	100%	Percent of instances where search data was accurately verified and fused with prior data resulting in correct determination	90%	100%	95%	1	
			33%	Timeliness	100%	Time for search data to be verified and fused with prior data which result in a correct determination	10 min	5 min	12 min	0	
			34%	Information Reliability	100%	Percent of instances where reliability of the data was maintained or improved from the verify and fuse process	90%	95%	90%	1	
Authenticate IP	50%	1	50%	Accuracy	100%	Percent of instances where the authentication of the IP to the location was accurate with real truth information	90%	95%	90%	1	1.0
			50%	Timeliness	100%	Time to authenticate IP from time first located	10 min	5 min	9 min	1	
	50%	2	50%	Accuracy	100%	Percent of instances where the authentication of the IP to the location was accurate with real truth information	90%	95%	90%	1	
			50%	Timeliness	100%	Time to authenticate IP from time first located	10 min	5 min	9 min	1	

- **Measuring Risk.** Each measure is assessed for risk of error in determining if it met the threshold value. Table 3-22 provides the example for the JPR tasks supported by the PLB SUT. Column 13 is the likelihood of error based on the data analysis and table 2-11. Column 14 is the calculated risk for each attribute, and column 15 is the risk score (Col 15 ÷ 25).

**Table 3-22. Task Performance Risk Scoring Table**

Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 12	Col 13	Col 14	Col 15
Task	Condition Weight (W <sub>c</sub> )	Condition Set	Attribute Weight (W <sub>a</sub> )	Attribute	Measure Weight (W <sub>m</sub> )	Measure	Total Score (S <sub>i</sub> )	Likelihood of Error (L <sub>cam</sub> )	Calc. Risk	Risk Score (R <sub>i</sub> )
Execute Search Plan	50%	1	33%	Accuracy	100%	Percent of instances where search was executed according to plan before the IP was found	0.66	3	1.99	0.08
			33%	Timeliness	100%	Time to commence search from time tasked		2		
			34%	Information Reliability	100%	Count of search passes over IP position before IP located		1		
	50%	2	33%	Accuracy	100%	Percent of instances where search was executed according to plan before the IP was found		3		
			33%	Timeliness	100%	Time to commence search from time tasked		2		
			34%	Information Reliability	100%	Count of search passes over IP position before IP located		1		
Verify/Fuse Location	50%	1	33%	Accuracy	100%	Percent of instances where search data was accurately verified and fused with prior data resulting in correct determination	0.84	2	2.01	0.08
			33%	Timeliness	100%	Time for search data to be verified and fused with prior data which result in a correct determination		1		
			34%	Information Reliability	100%	Percent of instances where reliability of the data was maintained or improved from the verify and fuse process		3		
	50%	2	33%	Accuracy	100%	Percent of instances where search data was accurately verified and fused with prior data resulting in correct determination		2		
			33%	Timeliness	100%	Time for search data to be verified and fused with prior data which result in a correct determination		1		
			34%	Information Reliability	100%	Percent of instances where reliability of the data was maintained or improved from the verify and fuse process		3		
Authenticate IP	50%	1	50%	Accuracy	100%	Percent of instances where the authentication of the IP to the location was accurate with real truth information	1.0	2	1.5	0.06
			50%	Timeliness	100%	Time to authenticate IP from time first located		1		
	50%	2	50%	Accuracy	100%	Percent of instances where the authentication of the IP to the location was accurate with real truth information		2		
			50%	Timeliness	100%	Time to authenticate IP from time first located		1		

- **Aggregating Scores.** Aggregated scores for task performance attributes may be determined for both measures and risk. They are based on the weights assigned to the tasks found in table 3-9. Aggregating scores allows the decision-maker to see a single value that can be used as a gauge for SUT impact on task performance.
  - **Aggregate Task Performance Measures.** Tables 3-9 and 3-21 provide the information needed to aggregate task attribute measures into one score. The calculated aggregate score is shown in table 3-23. Based on the color codes for scores in table 2-9, the aggregated score for task performance is green.

**Table 3-23. Aggregate Task Performance Scoring Table**

Table 3-9	Table 3-21		New Column
Column 5	Column 1	Column 12	
Calc. Task Weight ( $W_t$ )	Task	Score ( $S_t$ )	Aggregate Score ( $S_{task}$ )
0.333	Execute Search Plan	0.66	0.833
0.333	Verify/Fuse Location	0.84	
0.333	Authenticate IP	1.00	

- **Aggregate Task Performance Risk.** The process for aggregating task performance risk in the assessment again uses information from tables 3-9 and 3-22. The calculated aggregate risk score is shown in table 3-24. Based on the color codes for scores in table 2-10, the aggregated risk score for task performance is green.

**Table 3-24. Aggregate Task Performance Risk Scoring Table**

Table 3-9	Table 3-21		New Column
Column 5	Column1	Column 12	
Calc. Task Weight ( $W_t$ )	Task	Risk Score ( $R_t$ )	Aggregate Score ( $R_{task}$ )
0.333	Execute Search Plan	0.08	0.073
0.333	Verify/Fuse Location	0.08	
0.333	Authenticate IP	0.06	

- **Identifying Exceptions.** The exceptions report shown in table 3-25 helps to explain deficiencies with task performance attribute scores and risk.

**Table 3-25. Task Performance Exceptions Report**

Column 1	Column 2	Column 3	Column 4
<b>Task Attribute Score Exceptions</b>			
ID No.	Task	Attribute	Exception
E9	Execute Search Plan	Information Reliability	Several instances where over IP electronic position, but could not visually identify
E10	Execute Search Plan	Information Reliability	More difficult visually locating IP in tropical terrain
E11	Verify/Fuse Location	Timeliness	Difficulty fusing data with incomplete information
E12	Verify/Fuse Location	Information Reliability	Weak signal impacted ability to verify and fuse data
<b>Task Attribute Risk Exceptions</b>			
ID No.	Task	Attribute	Exception
E13	Verify/Fuse Location	Information Reliability	Subjective assessment in determination of maintained or improved data

## Mission Effectiveness Assessment

### Overview

This example will use a qualitative assessment of mission effectiveness based on the findings at the system/SoS and task performance levels.

### Process

The qualitative assessment of mission effectiveness is based on COIs for each mission and sub-COIs for each mission desired effect. For a JPR mission, they are shown in table 3-26. The question is whether the assessment is required on all four desired effects.

**Table 3-26. Mission COI and Sub-COIs**

<b>Mission</b>	JPR	<b>COI:</b> Can PBR SUT support the JPR mission?
<b>Desired Effect</b>	Return isolated personnel to duty	<b>Sub-COI:</b> Assess the ability to return IP to duty
<b>Desired Effect</b>	Sustain morale	<b>Sub-COI:</b> Assess the ability to maintain morale of the IP
<b>Desired Effect</b>	Increase operational performance	<b>Sub-COI:</b> Assess the ability to increase operational performance
<b>Desired Effect</b>	Deny adversary ability to exploit the IP	<b>Sub-COI:</b> Assess the ability to deny the adversary an ability to exploit the IP

It was determined earlier that the PLB SUT only supported three sub-tasks under the “Locate” task in the JPR mission thread. Using tables 3-10, 3-3, and 3-7 (in that order), those three sub-tasks map to three of the four desired effects for the JPR mission. The relationships are shown and high-lighted in table 3-27. Therefore, in performing our qualitative assessment of the mission, the first desired effect does not have to be addressed.

**Table 3-27. Tasks to Mission Desired Effects Relationships**

Tasks	Execute Search Plan	Verify/Fuse Location	Authenticate IP	Share Location	↓ Table 3-10						↓ Table 3-3				
					Task - Attributes										
	X	X	X	X	Accuracy			X							
	X	X	X	X	Timeliness			X							
	X	X			Information Reliability					X	X				
				X	Completeness		X				X				
Table 3-7 →					Mission Desired Effects	Mission Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness				
					Return isolated personnel to duty		X								
					Sustain morale			X							
					Increase operational performance				X	X					
					Deny adversary opportunity to exploit the IP					X	X				

The assessment of the COI and sub-COIs is based on subjective insights from the system/SoS and task performance. Table 3-28 provides an abbreviated assessment of the PLB SUT impact on mission effectiveness.

**Table 3-28. Mission Assessment**

<b>Mission COI:</b> Can PBR SUT support the JPR mission?		
	<b>Sub-COI:</b> Assess the ability to maintain morale of the IP	
	Responsiveness:	
		Time delays that may impact the morale of the IP
		Difficulties maintaining contact with the IP may impact morale
		Difficulties in relaying status updates to IP may impact morale
	<b>Sub-COI:</b> Assess the ability to increase operational performance	
	Coordination	
		Failures to maintain electronic positions of IP will impact operational performance
	Awareness	
		Maintaining of conflicting data will hinder verify/fuse of data and operational performance
	<b>Sub-COI:</b> Assess the ability to deny the adversary an ability to exploit the IP	
	Awareness	
		Maintaining established communications with IP will help deny adversary exploitation

## STEP 3: REPORTING STANDARD

An example report of the PLB SUT effectiveness and suitability, to include impacts on task performance and mission effectiveness, is provided in this section.

**System/SoS Assessment Reporting.** The PLB system/SoS assessment report is shown as table 3-29. It may be concluded that:

- The “Operational Capability” KPP passed, but there is some risk in the assessment. Additional testing to collect more data may be desired.
- The “Net-Ready” KPP failed due to inability to operate with the P-3 airframe. Modifications may be necessary to pass this KPP.
- The system failed shock resistance testing, and the system may require modifications to meet the threshold value.
- Life-cycle maintenance costs are estimated to be within the threshold value.

**Table 3-29. PLB System/SoS Assessment Report**

SECTION 1					
Attribute Category	Score		Risk Score	Exceptions	
Effectiveness KPPs					
Operational Capability	1.00		0.60	E6	
Net-Ready	0.0		0.20	E1	
Suitability KPPs					
SECTION 2					
Attribute Category	Count	Count Passed	Count Failed	Avg Risk Score	Exceptions
Effectiveness Attributes					
KSAs	2	2	0	0.15	E2, E4
OAs	2	1	1	0.10	E5
Suitability Attributes					
KSAs	1	1	0	0.45	E3, E7
OAs	2	2	0	0.14	E8
SECTION 3					
Exceptions - Effectiveness					
ID No.	Attribute	Type	Exception		
E1	Net-Ready	KPP	Not interoperable with P-3		
E2	Protection	KSA	Some data packets dropped over time		
E4	Interoperability	KSA	Power output affected in cold weather		
E5	Shock Resistant	OA	Failure after several shock tests		
E6	Operational Capability	KPP	Limited data not statistically significant		
Exceptions - Suitability					
ID No.	Attribute	Type	Exception		
E3	Sustainment	KSA	Reliability affected in cold weather		
E7	Sustainment	KSA	Maintenance costs estimated over time		
E8	Battery Life	OA	Unable to fully test		

**Task Performance Reporting.** The PLB system/SoS assessment report of impacts on task performance is shown as table 3-30. It may be concluded that:

- Incomplete data from PLB caused confusion and difficulties in verify/fuse IP location.
- Difficulty visually locating IP in tropical terrain, even with accurate PLB position data.
- Subjective assessment of verify/fuse IP location impacts risk in assessment.

**Table 3-30. PLB Impacts on Task Performance Assessment Report**

SECTION 1					
Overall Mission	Score		Risk Score	Exceptions	
Mission: JPR Tasks	0.833		0.073	E9 – E13	
Mission: N/A					
SECTION 2					
Tasks	Count	Ct Passed	Ct Failed	Avg Risk Score	Exceptions
Mission: JPR					
Tasks performed by SUT	0	0	0		
Tasks supported by SUT	3	3	0	0.073	E9 – E13
Mission: N/A					
Tasks performed by SUT					
Tasks supported by SUT					
SECTION 3					
Mission One					
ID No.	Task	Attribute	Exception		
E9	Execute Search Plan	Information Reliability	Several instances where over IP electronic position but could not visually ID		
E10	Execute Search Plan	Information Reliability	More difficult visually locating IP in tropical terrain		
E11	Verify/Fuse Location	Timeliness	Difficulty fusing data with incomplete information		
E12	Verify/Fuse Location	Information Reliability	Weak signal impacted ability to verify and fuse data		
E13	Verify/Fuse Location	Information Reliability	Subjective assessment in determination of maintained or improved data		

**Mission Effectiveness Reporting.** Table 3-31 (repeat of table 3-28) shows the qualitative assessment of the PLB impacts on mission effectiveness. It may be concluded that:

- Incomplete data from PLB caused time delays and reliability issues that can impact IP morale.
- Inability to maintain electronic position data can negatively impact operational performance.
- Maintaining communications with the IP will help to avoid exploitation by the adversary.



**Table 3-31. Mission Assessment**

<b>Mission COI:</b> Can PBR SUT support the JPR mission?		
	<b>Sub-COI:</b> Assess the ability to maintain morale of the IP	
	Responsiveness	
	Time delays that may impact the morale of the IP	
	Difficulties maintaining contact with the IP may impact morale	
	Difficulties in relaying status updates to IP may impact morale	
	<b>Sub-COI:</b> Assess the ability to increase operational performance	
	Coordination	
	Failures to maintain electronic positions of IP will impact operational performance	
	Awareness	
	Maintaining of conflicting data will hinder verify/fuse of data and operational performance	
	<b>Sub-COI:</b> Assess the ability to deny the adversary an ability to exploit the IP	
	Awareness	
	Maintaining established communications with IP will help deny adversary exploitation	

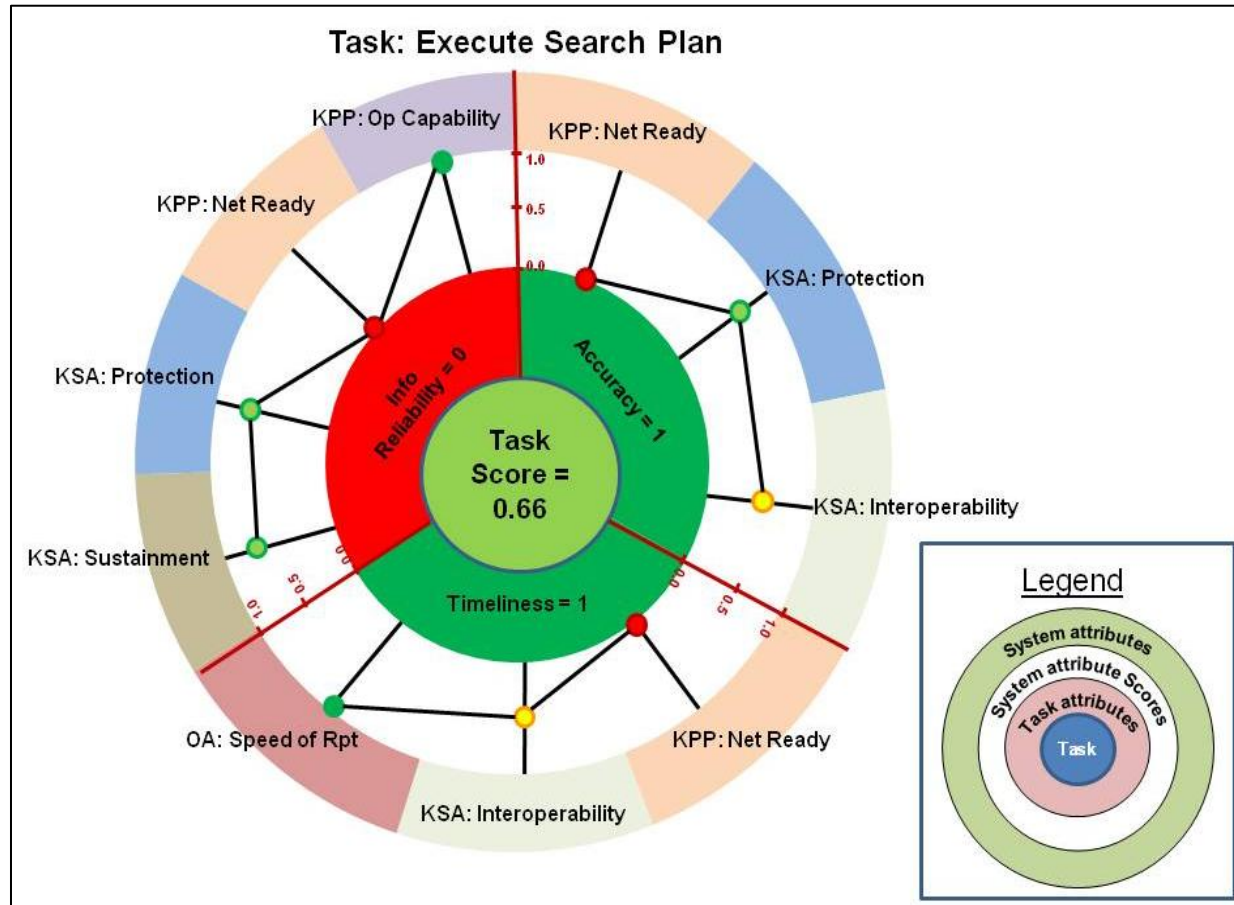
**Root Cause Analysis.** A root cause analysis is conducted to determine how the system impacted task performance and mission desired effects. Table 3-32 shows the linkages between system attributes, task attributes, and mission effects. The information is derived from other constructed tables (3-16, 3-21, 3-27, and C-8). It can be seen that the “execute search plan” task did not score as high as desired, primarily due to its “information reliability” attribute. A further look at cause from the system appears that the failure of the Net-Ready KPP was a primary contributor for the degradation in the task attribute. However, that same system KPP had little impact on the “timeliness” and “accuracy” attributes for the same task.

Figure 3-1 provides an alternative means to display the causal relationships of system attributes to the task “execute search plan.” In addition to the information found in table 3-32, the diagram shows the weighting of system and task attributes to gain an understanding of the relative impacts the system has on the task performance. In this example, each system attribute was equally weighted for each task attribute.

It can also be seen in table 3-32 that the “timeliness” attribute of the “verify and fuse” task is affected by the system Net-Ready KPP and Interoperability KSA. Measuring the task showed that time was impacted due to incomplete information and/or the loss of information. The causal relationships showed that the system network issues and reliability of the PLB power was the reason for incomplete information. Table 3-33 summarizes the root cause analysis of the PLB impacts on task performance and mission effects.

Table 3-32. Mission-Task-System Linkages

Mission	Desired Effects (Sub-COIs)	Tasks	Task Score	Task Attributes	Task Attribute Score	System Attributes	System Attribute Score
Joint Personnel Recovery	(1) Assess the ability to return isolated personnel to duty (2) Assess the ability to maintain morale (3) Assess the ability to increase operational performance (4) Assess the ability to deny adversary opportunity to exploit the IP	Execute Search Plan	0.66	Accuracy	1.0	KPP: Net-Ready	0.0
						KSA: Protection	0.75
						KSA: Interoperability	0.5
				Timeliness	1.0	KPP: Net-Ready	0.0
						KSA: Interoperability	0.5
						OA: Speed of Initial Report	1.0
				Information Reliability	0.0	KPP: Operational Capability	1.0
						KPP: Net-Ready	0.0
						KSA: Protection	0.75
		Verify/Fuse Location	0.84	Accuracy	1.0	KSA: Sustainment	0.75
						KPP: Net-Ready	0.0
						KSA: Protection	0.75
				Timeliness	0.5	KSA: Interoperability	0.5
						KPP: Net-Ready	0.0
						KSA: Interoperability	0.5
				Information Reliability	1.0	OA: Speed of Initial Report	1.0
						KPP: Operational Capability	1.0
						KPP: Net-Ready	0.0
		Authenticate IP	1.0	Accuracy	1.0	KSA: Protection	0.75
						KPP: Net-Ready	0.0
				Completeness	1.0	KSA: Protection	0.75
						KSA: Sustainment	0.75



**Figure 3-1. System - Task Causality Diagram**

**Table 3-33. Root Cause Analysis**

System/SoS Functions			Task Performance				Mission Effects	
Attribute	Issue		Attribute	Task	Issue		Attribute	Desired Effects
Net-Ready	Inability to network with P-3 A/C	Degraded →	Information reliability	Execute search plan	Inability to visually ID	Degraded →	Coordination	Ability to increase operational performance
Protection	Some data packets dropped over time	Degraded →	Information reliability	Execute search plan	Inability to visually ID	Degraded →	Coordination	Ability to increase operational performance
Sustainment	Inability to maintain power output in cold weather	Degraded →	Information reliability	Execute search plan	Inability to visually ID	Degraded →	Responsiveness	Ability to sustain morale of the IP
Net-Ready	Inability to network with P-3 A/C	Degraded →	Timeliness	Verify/fuse data	Incomplete information	Degraded →	Coordination	Ability to increase operational performance
Sustainment	Reliability of the PLB	Degraded →	Timeliness	Verify/fuse data	Weak signal caused loss of information	Degraded →	Awareness	Ability to deny adversary opportunity to exploit the IP

## CHAPTER 4 SUMMARY

### INTRODUCTION

The Measures Development SOP begins the mission-based T&E process by decomposing the mission to tasks to needed system functions and develops measures for each. The process follows the left side of the T&E-V diagram shown in figure 1-1. This guidebook completes the right side of the T&E-V diagram by conducting an assessment of system impact on the SoS ability to perform tasks and achieve mission desired effects.

### FOCUS

The intent of the acquisition process is to support the warfighter by filling identified capability gaps. Capability gaps are described in terms of the risk to mission (the ability to achieve the objectives of the scenario), the risk to force (the potential losses due to the capability gap), and other important considerations, such as resourcing risks and affects on allies.<sup>11</sup> Additionally, JCIDS indicates capability gaps are characterized as to whether they are due to:

- Proficiency (ability to achieve the relevant effect in particular conditions)
- Sufficiency (ability to achieve the effect but inability to bring the needed force to bear due to force shortages or other commitments)
- Lack of existing capability
- Need for replacement due to aging of an existing capability
- Policy limitations (inability to use the force as needed due to policy constraints)

The focus of the T&E-V process, documented in the Measures Development SOP and this guidebook, is to assess the filling of capability gaps through an evaluation of mission effects and task performance. It places the focus on evaluating the impact on the warfighter, not just on a system's ability to function.

### Example

A radio frequency jammer is designed to jam frequencies of common household devices (garage door openers, remote controls, and so forth) that could be used to remotely detonate an Improvised Explosive Device (IED). The system is mounted on a Humvee and found to jam the proper frequencies at a proper range. However, the impacts on mission were never assessed.

After fielding the jammer, it was found that when turned on the system hindered the ability of vehicles in the convoy to communicate with each other. This impacted the speed of the convoy and ability to maintain contact with the command. Thus, task performance was degraded and mission effects not achieved.

---

<sup>11</sup> JCIDS Manual, Page A-6, Dated July 2009

## **FLEXIBILITY**

The assessment process offers a great amount of flexibility to the evaluator through the use of various assessment models. Various scoring models are discussed in chapter 1 of this guidebook and described in annex D. The assessor has the option to select the model that best suits their needs. Other models in the assessment process also provide flexibility. The risk model, color coding models, and weighting criteria can all be adjusted to suit the needs of the user. It may be found that a commander of a test organization may, over time, dictate what models to use and the standards for each model. The risk assessment is based on the likelihood of an error and impact if there is an error. Likelihood is broken down into reliability and validity. Test commanders may wish to add more definition to the process based on their acceptable levels of risk.

## **ENABLING DESIGN OF EXPERIMENTS**

### **Quote**

*“The experimenter who does not know what he is looking for will not understand what he finds.”* Claude Bernard

This T&E-V process enables the design of experiment process to focus on mission and task attributes and measures that are important to the warfighter and that provide evidence the capability gaps have been filled. It establishes a robust and repeatable process to determine dependent and independent variables that are cause and effect relationships needed in the design. It supports the identification of conditions, scenarios and vignettes, data requirements, test methods, and resource requirements

## ANNEX A

### ACRONYMS AND ABBREVIATIONS

ACRONYM OR ABBREVIATION	DEFINITION
ATEC	Army Test and Evaluation Command
C2	Command and Control
CAS	Close Air Support
CBA	Capabilities-Based Assessment
CDD	Capability Development Document
CJCSI	Chairman of the Joint Chiefs of Staff Instruction
CJCSM	Chairman of the Joint Chiefs of Staff Manual
COI	Critical Operational Issue
CONOPS	Concept of Operations
CTM	Capability Test Methodology
CTP	Critical Technical Parameter
DE	Desired Effect
DoD	Department of Defense
DoDAF	Department of Defense Architecture Framework
DODI	Department of Defense Instruction
DOE	Design of Experiments
DOT&E	Director, Operational Test and Evaluation
DOTMLPF	Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, and Facilities
ICD	Initial Capabilities Document
IDEF0	Integration Definition Model
IP	Isolated Personnel
ISOPREP	Isolated Personnel Report
JCA	Joint Capability Area
JCIDS	Joint Capabilities Integration and Development System
JMT	Joint Mission Thread
JP	Joint Publication
JPR	Joint Personnel Recovery
JT&E	Joint Test and Evaluation
JTEM	Joint Test and Evaluation Methodology
JTEM-T	Joint Test and Evaluation Methodology - Transition
KPP	Key Performance Parameter
KSA	Key System Attribute
MBT&E	Mission-Based Test and Evaluation
MLM	Mission Level Measure
MOE	Measure of Effectiveness
MOP	Measure of Performance
NATO	North Atlantic Treaty Organization

ACRONYM OR ABBREVIATION	DEFINITION
OA	Other Attributes
OPLAN	Operational Plan
OT&E	Operational Test and Evaluation
OV	Operational Viewpoint
PLB	Personal Locator Beacon
SME	Subject Matter Expert
SOP	Standard Operating Procedure
SoS	System-of-Systems
SUT	System Under Test
T&E	Test and Evaluation
T&E-V	JTEM-T Test and Evaluation “V” Diagram (Figure 1-1)



## ANNEX B

### DEFINITIONS OF TERMS

The following terms of reference establish a lexicon for discussing measures development. Whenever possible, definitions were taken from authoritative joint publications.

**Activity:** An activity is work not specific to a single organization, weapon system, or individual that transforms inputs into outputs or changes their state. (Department of Defense Architecture Framework [DoDAF] Version 2.0)

**Attribute:** A quantitative or qualitative characteristic of an element or its actions. (*Manual for the Operation of the Joint Capabilities Integration and Development System* [JCIDS Manual], Revised July 31, 2009, and Chairman, Joint Chiefs of Staff Manual [CJCSM] 3170.01C, Cancelled)

**Capability:** The ability to achieve a desired effect under specified standards and conditions through combinations of means and ways across doctrine, organization, training, materiel, leadership and education, personnel, and facilities (DOTMLPF) to perform a set of tasks to execute a specified course of action. (JCIDS Manual, Revised July 31, 2009)

**Condition:**

1. Those variables of an operational environment or situation in which a unit, system, or individual is expected to operate and may affect performance. (CJCSM 3500.04E, *Universal Joint Task Manual*, August 25, 2008)
2. The sample of adversaries and operating conditions – the scenario. (*Capabilities-Based Assessment User's Guide*, Version 3, March 2009)

**Criterion:** The minimum acceptable level of performance associated with a particular measure of (task) performance. It is often expressed as hours, days, percent, occurrences, minutes, miles, or some other command-stated measure. (CJCSM 3500.04E, *Universal Joint Task Manual*, August 25, 2008)

**Effect (Mission Desired):**

1. The physical or behavioral state of a system that results from an action, a set of actions, or another effect.
  2. The result, outcome, or consequence of an action.
  3. A change to a condition, behavior, or degree of freedom.
- (Joint Publication [JP] 1-02, *Department of Defense Dictionary of Military and Associated Terms*, April 12, 2001, as amended September 2010)

**Function (System/Operational):** The action for which a person or thing is specially designed, fitted, used, or intended to accomplish or execute. (DoDAF 2.0)

**Joint Mission Environment:** A subset of the joint operational environment composed of force and non-force entities and conditions, circumstances, and influences within which forces employ

capabilities to execute joint tasks to meet a specific mission objective. (JCIDS Manual, Revised July 31, 2009)

**Key Performance Parameter (KPP)/Key System Attribute (KSA)/Critical Technical Parameter (CTP):** Attributes and/or parameters of a system that are considered critical. (JCIDS Manual, Revised July 31, 2009)

**Means:**

1. Forces, units, equipment, and resources. (*Terms of Reference [TOR] for Conducting a Joint Capability Area (JCA) Baseline Reassessment*, April 9, 2007)
2. Solutions represent means or resources that can be employed. (*Capabilities-Based Assessment User's Guide*, Version 3, March 2009)
3. Means are based on DOTMLPF organization, materiel, personnel, and facility resources.

**Measure:** A parameter that provides the basis for describing varying levels of task accomplishment. (CJCSM 3500.04E, *Universal Joint Task Manual*, August 2008)

**Measure of Effectiveness:** A criterion used to assess changes in system behavior, capability, or operational environment that is tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect. (JP 1-02, April 12, 2001, as amended September 2010)

**Measure of Performance:** A criterion used to assess friendly actions that are tied to measuring task accomplishment. (JP 1-02, April 12, 2001, as amended September 2010)

**Measure of Suitability:** A measure of an item's ability to be supported in its intended operational environment. (Defense Acquisition University Glossary, 13th Edition, November 2009)

**Measure of System/System of Systems (SoS) Attribute:** A parameter that describes varying levels of attributes. (Capability Test Methodology Handbooks, April 2009)

**Mission:** The task, together with the purpose, that clearly indicates the action to be taken and the reason therefore. (JP 1-02, April 12, 2001, as amended September 2010)

**Node:** An element of a system that represents a person, place, or physical thing. (JP 1-02, April 12, 2001, as amended September 2010)

**Objective Value:** The desired operational goal associated with a performance attribute beyond which any gain in utility does not warrant additional expenditure. The objective value is an operationally significant increment above the threshold. An objective value may be the same as the threshold when an operationally significant increment above the threshold is not significant or useful. (JCIDS Manual, July 2009)

**Reliability:** The extent to which the measure produces the same result when used repeatedly to measure the same thing. (Rossi, Peter H., Lipsey, Mark W., and Freeman, Howard E.,

*Evaluation: A Systematic Approach*, Seventh Edition, Sage Publications, Inc.: Thousand Oaks, CA, 2004)

**Sensitivity:** The extent to which the values of the measure change when a change or difference occurs in the thing being measured. (Same source as “Reliability”)

**Standard:** A standard provides a way of expressing the acceptable proficiency that a joint organization or force must perform under a specified set of conditions. A standard consists of one or more measures for a task and a criterion for each measure. (CJCSM 3500.04E, *Universal Joint Task Manual*, August 2008)

**System:** A functionally, physically, and/or behaviorally related group of regularly interacting or interdependent elements; that group of elements forming a unified whole. (JP 1-02, April 12, 2001, as amended September 2010)

**System-of-Systems (SoS):** A set or arrangement that results when independent and useful systems are integrated into a larger system that delivers unique capabilities. (JCIDS Manual, Revised July 31, 2009)

**Task:** An action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability. (CJCSM 3500.04E, *Universal Joint Task Manual*, August 2008) NOTE: This term and its definition are to be included in JP 1-02.

**Threshold Value:** A minimum acceptable operational value below which the utility of the system becomes questionable. (JCIDS Manual, July 2009)

**Validity:** The extent to which the measure succeeds at measuring what it is intended to measure. (Same source as “Reliability”)

**Ways:**

1. Doctrine; tactics, techniques, and procedures; competencies; and concepts. (*TOR for Conducting a JCA Baseline Reassessment*, April 9, 2007)
2. Functions [are] considered ways. (*Capabilities-Based Assessment User’s Guide*, Version 3, March 2009)
3. Ways are based on DOTMLPF doctrine, training, and leadership.

This page intentionally left blank.

## ANNEX C

### MEASURES DEVELOPMENT EXAMPLE MATRICES

The following figures are published in the Measures Development Standard Operating Procedure and are captured in this annex for reference purposes only.

<b>Mission statement:</b>					
Prevent, prepare, plan, execute, and adapt military, diplomatic, and civil efforts to conduct Joint Personnel Recovery for isolated personnel who are separated (as an individual or group) from their unit while participating in a US-sponsored military activity or mission and who are, or may be, in a situation where they must survive, evade, resist, or escape in order to recover and reintegrate those isolated personnel into their organization. (JP3-50)					
<b>Objectives</b>	<b>Desired Effects</b>	Return isolated personnel to duty (JP3-50, pg ix)	Sustain morale (JP3-50, pg ix)	Increase operational performance (JP3-50, pg ix)	Deny adversary the opportunity to influence our military strategy and national will by exploiting the intelligence and propaganda value of isolated personnel (JP3-50, pg ix)
Protect the force (JPR JMT Annex 2, pg 82)		X	X		
Enable military missions (JPR JMT Annex 2, pg 82)		X	X	X	
Defeat adversary attempts to exploit a known asymmetric vulnerability (JPR JMT Annex 2, pg 82)		X			X

**Figure C-1. Matrix 1 Example –  
Joint Personnel Recovery (JPR) Joint Mission Thread (JMT)**

#	Desired Effects	Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness
DE1	Return isolated personnel to duty (JP3-50, pg ix)		X				
DE2	Sustain morale (JP3-50, pg ix)			X			
DE3	Increase operational performance				X	X	
DE4	Deny adversary opportunity to exploit the IP (JP3-50, pg ix)				X		X

Figure C-2. Matrix 2 Example – JPR JMT

#	Scale	Mission Measures (MM)	Attributes	Availability	Responsiveness	Coordination	Awareness	Readiness
MM-1	Percent	Of JPR missions where IP personnel was available to be cleared for duty		X				
MM-2	Percent	Of JPR missions where IP personnel was cleared for duty		X				
MM-3	Time	For IP to respond to changes in threat and environmental conditions that required the IP to evade, resist, or escape			X			
MM-4	Percent	Of JPR missions where morale was a factor in IP inability to survive, evade, resist, or escape			X			
MM-5	Percent	Of JPR missions where no unplanned redundant activities occurred				X		
MM-6	Percent	Of JPR missions where continuous horizontal coordination existed across operational nodes				X		
MM-7	Percent	Of JPR missions where continuous vertical coordination existed across operational nodes				X		
MM-8	Percent	Of JPR mission executions where planning and preparation led to successful coordination across operational nodes				X		
MM-9	Percent	Of JPR missions where correct decisions were made by operational nodes based on situational awareness					X	
MM-10	Percent	Of JPR missions where IP acted correctly based on situational awareness					X	
MM-11	Percent	Of JPR missions where operational nodes acted correctly based on situational awareness					X	
MM-12	Percent	Of JPR missions where inadequate training led to mission execution deficiencies						X
MM-13	Percent	Of JPR missions where inadequate systems, supplies, and resources led to mission execution deficiencies						X

Figure C-3. Matrix 3 Example – JPR JMT

<b>Attribute</b>	<b>CCI #</b>	<b>Critical Capability Issue (mission)</b>	<b>Measure #</b>
<b>Availability</b>	<b>1</b>	<b>Assess the ability to recover and clear the IP for duty</b>	MM-1
			MM-2
<b>Responsiveness</b>	<b>2</b>	<b>Assess the ability to maintain morale of the IP</b>	MM-3
			MM-4
<b>Coordination</b>	<b>3</b>	<b>Assess the ability to coordinate JPR missions across operational nodes that reduces redundancy and creates synergy</b>	MM-5
			MM-6
			MM-7
			MM-8
<b>Awareness</b>	<b>4</b>	<b>Assess the ability to maintain situational awareness across all operational nodes</b>	MM-9
			MM-10
			MM-11
<b>Readiness</b>	<b>5</b>	<b>Assess the ability to maintain a state of preparedness for conducting JPR missions</b>	MM-12
			MM-13

**Figure C-4. Mission Critical Issues – JPR JMT**

<b>Task #</b>	<b>Operational Tasks/Activities</b>	<b>Joint Personnel Recovery Center (JPRC)</b>	<b>Personnel Recovery Coordination Center (PRCC)</b>	<b>Battle Coordination Detachment</b>	<b>Search Team</b>	<b>Defense POW/Mission Personnel Office (DPMO)</b>
4.0	Locate	X	X	X	X	X
4.1	Execute Search Plan	X			X	X
4.1.1	Continue to Monitor and Update Location				X	
4.1.2	Utilize Information Source	X			X	
4.1.3	Search for Location				X	
4.1.4	Apply Objective Area Search	X			X	
4.1.5	Apply Area and Pattern Search	X			X	
4.1.6	Apply Visual Search	X			X	
4.1.7	Apply Electronic Search	X			X	
4.1.8	Process Search Location	X				X
4.2	Verify/Fuse Location	X		X		
4.3	Authenticate IP	X		X	X	
4.4	Share Location	X	X			

**Figure C-5. Matrix 4 Example – JPR JMT**



Task #	<u>Operational Tasks/Activities</u>		<u>Attributes</u>	Accuracy	Timeliness	Information Reliability	Completeness
4.0	Locate			X	X	X	X
4.1	Execute Search Plan	Determine the location and status of the isolated personnel (precisely find; fast response)		X	X	X	
4.2	Verify/Fuse Location	Verify and fuse isolated personnel's location information to provide accurate and reliable coordinates for refining recovery plans. Goal is for latest, most reliable location information.		X	X	X	
4.3	Authenticate IP	Authenticate isolated personnel using ISOPREP data and other methods		X			X
4.4	Share Location	Use available information to refine isolated personnel's location with reliable and accurate information		X	X		X

Figure C-6. Matrix 5 Example – JPR JMT



Task #	Operational Tasks/Activities	Attributes	Accuracy	Timeliness	Information Reliability	Completeness	Task Measure Number (TM-#)	Scale	Measure
4.0	Locate		X	X	X	X			
						X	TM-1	Percent	Of Locate tasks that located the IP
				X			TM-2	Time	To locate IP
			X				TM-3	Percent	Of instances where shared information on IP location was accurate with ground truth
						X	TM-4	Percent	Of Locate tasks where information shared was complete
				X			TM-5	Percent	Of Locate tasks executed in a timely manner that provided an accurate location
					X		TM-6	Percent	Of Locate tasks that maintained or improved information reliability of the shared information
4.1	Execute Search Plan			X					
			X				TM-8	Percent	Of instances where search was executed according to plan before the IP was found
				X			TM-9	Time	To commence search from time tasked
					X		TM-10	Count	Of search passes over IP position before IP located
4.2	Verify/Fuse Location		X	X	X				
			X				TM-11	Percent	Of instances where search data was accurately verified and fused with prior data resulting in correct determination
				X			TM-12	Time	For search data to be verified and fused with prior data which result in a correct determination
					X		TM-13	Percent	Of instances where reliability of the data was maintained or improved from the verify and fuse process
4.3	Authenticate IP		X	X					
			X				TM-14	Percent	Of instances where the authentication of the IP to the location was accurate with real truth information
				X			TM-15	Time	To authenticate IP from time first located
4.4	Share Location		X	X		X			
			X				TM-16	Percent	Of shared information exchanges where information received was accurate with what was sent
						X	TM-17	Percent	Of shared information exchanges where information received was complete when compared to what was sent
				X			TM-18	Time	To execute the share location sub-task where information exchange was accurate and complete

Figure C-7. Matrix 6 Example - JPR JMT

JCA	KPP	KPP Attribute	Threshold	Objective
C2	Operational Capability	Signal range	50 NM range	100 NM range
C2	Net-Ready	Interoperable	Interoperable with 100% U.S. SAR systems	Interoperable with 100% U.S. & coalition SAR systems
JCA	KSA	KSA Attribute	Threshold	Objective
Force protection	Protection	Transmitted data accuracy	99% data accuracy	Same as threshold
Force protection	Protection	Access and control	Single hand controllable	Same as threshold
Logistics	Sustainment	Reliability	95% probability operational for 24 hr period	99% probability operational for 24 hr period
Logistics	Sustainment	Ownership cost	\$50 annual upkeep cost	\$25 annual upkeep cost
C2	Interoperability	Transmission output	20 watt continuous power	25 watt continuous power
		Other System Attributes	Threshold	Objective
		Shock resistant	Withstand ejection seat shock	Same as threshold
		Speed of initial report	5 sec after activation	2 sec after activation
		Water resistant	Watertight to 5m	Watertight to 10m
		Battery life	5 year	7 year

Figure C-8. SUT Attributes (Notional)

Task		Locate											
Sub-task		Execute search plan			Verify & fuse location			Authenti- cate IP		Share location			
System/SoS Attribute		Task Attribute	Accuracy	Timeliness	Info Reliability	Accuracy	Timeliness	Info Reliability	Accuracy	Completeness	Accuracy	Timeliness	Completeness
Type	Attribute												
KPP	Operational Capability: Signal range			X			X						
KPP	Net-Ready: Interoperable		X	X	X	X	X	X	X	X			
KSA	Protection: Transmitted data accuracy		X		X	X		X	X				
KSA	Protection: Access and control									X			
KSA	Sustainment: Reliability				X			X		X			
KSA	Sustainment: Ownership cost												
KSA	Interoperability: Transmission output		X	X		X	X						
OSA	Shock resistant												
OSA	Speed of initial report			X			X						
OSA	Water resistant												
OSA	Battery life												

Figure C-9. Matrix 7 Example – JPR JMT

System/SoS Attribute		System/SoS Attribute Measure		Conditions	
Type	Attribute	Scale	Measure Description	Condition	Descriptor
KPP	Operational capability: Signal range	NM	Max range for clear continuous signal	Ambient temperature	-20C
		NM	Max range for clear continuous signal	Ambient temperature	50C
KPP	Net-Ready: Interoperable	Pct	SAR systems interoperable with	Friendly forces	U.S. only
		Pct	SAR systems interoperable with	Friendly forces	U.S. and coalition
KSA	Protection: Transmitted data accuracy	Pct	Data transmissions that are complete		
		Pct	Data transmissions that are complete & accurate		
KSA	Protection: Access and control	Y/N	Single handed controllable operations		
KSA	Sustainment: Reliability	Pct	Probability operable for 24 hr period	Ambient temperature	-20C
		Pct	Probability operable for 24 hr period	Ambient temperature	50C
KSA	Sustainment: Ownership cost	\$\$\$	Annual maintenance cost		
KSA	Interoperability: Transmission output	Watts	Continuous transmission power output	Ambient temperature	-20C
		Watts	Continuous transmission power output	Ambient temperature	50C
OSA	Shock resistant	Pct	Operable after ejected from aircraft seat		
OSA	Speed of initial report	Sec	Time between activation & initial beacon broadcast		
OSA	Water resistant	Meters	Max depth maintains watertight		
OSA	Battery life	Years	Max battery shelf life		

**Figure C-10. Matrix 8 Example – JPR JMT**

This page intentionally left blank.



## ANNEX D

### ASSESSMENT PROCESS SCORING MODELS

Aggregating measures requires a common scale for the measures to prevent skewing of the results. Scoring models based on a measure's observed value compared to threshold and objective values provides a means to establish a common scale. This annex discusses a number of scoring models that may be used to aggregate measures. Each model has different strengths and weaknesses that must be considered in selecting the model to use. The assessment team should select a scoring model that will meet the needs of the assessment process.

This list of models is not all inclusive. Other models may be developed to better suit the needs of the assessment team. However, once a model is selected, the same model must be used for all the measures to ensure the scale remains constant.

The simplest model provided is the Threshold Model. This model is based on an ordinal scale (pass-fail) focused on whether the measure's observed value met the threshold value. The resulting score values are either 0 or 1. Other models are offered that have a continuous linear function based on current values, threshold values, and objective values (a current value is the measures value at the current capability). The examples in this annex are written in terms of tasks; however, the same models and functions can apply for measures of mission effectiveness and system under test (SUT)/system-of-systems (SoS) attributes.

1. **Threshold Model.** The Threshold Model is similar to a pass-fail model in that the result is based on whether the measure value met the threshold. Figure D-1 illustrates the score of the measure as a binary function of the threshold value with a score value of 0 or 1.

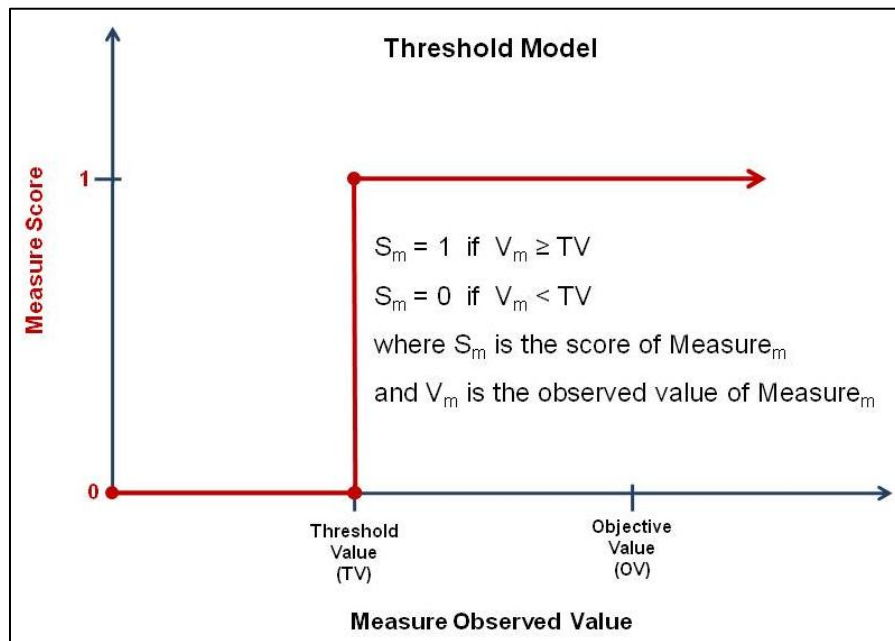


Figure D-1. Threshold Model

The Threshold Model scores can then be aggregated across measures to provide a single score. The aggregation can occur at any level (mission, task, and so forth). If assessing task performance (measures of performance [MOP]), then the measures are aggregated at the task level. If assessing mission effectiveness (measures of effectiveness [MOE]), then the measures are aggregated at the desired effect (DE) level. Assuming each attribute and measure have been properly weighted for impact on the task (mission), then the aggregate task score ( $S_t$ ) for a single task  $t$  can be given by:

The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission would be:

Pros:

- Provides a single value between 0 and 1
- Places emphasis on meeting threshold values
- If observed value is significantly higher than threshold value, then less stringent statistical requirements and less risk of error

Cons:

- Does not consider any incremental improvements above threshold value
- Does not add value to the score based on objective values

2. **Threshold – Weakest Link Model**. The Threshold – Weakest Link Model is the same as the Threshold Model in that each individual measure is evaluated using the same function shown in figure D-1. What differs is in the aggregation of those measures for all the tasks in a single mission. That is, the aggregation is based on the weakest measure for the task. If a single measure for the task fails to meet its threshold value, then the score for that task becomes 0. The aggregate task score ( $S_t$ ) can be given by:

The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is calculated the same as the Threshold Model, given as:

NOTE: The weak link could be applied to all the tasks in the mission, given as:

Pros:

- Provides a single value between 0 and 1
- Places increased emphasis on meeting threshold values
- Provides an all or nothing approach
- If observed value is significantly higher than threshold value, then less stringent statistical requirements and reduced risk

Cons:

- Does not consider any incremental improvements above threshold value
- Does not add value to the score based on objective values
- Might skew the scoring downward by assigning a 0 value to a task based on not meeting the threshold for a low priority (low weight) measure

3. **Threshold – Linear Model.** The Threshold – Linear Model considers the observed value of the measure to be compared to the threshold value as a ratio value. For example, an observed value half the threshold value would have a score of 1/2, whereas an observed value that is twice the value of the threshold would have a score of 2. Figure D-2 illustrates the score given to a measure based on the observed value and threshold value.

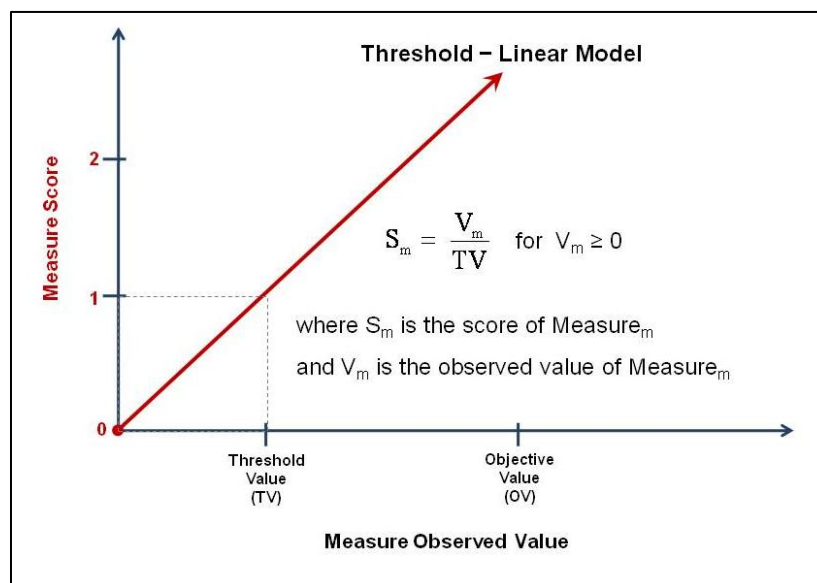


Figure D-2. Threshold - Linear Model

The TLM scores can then be aggregated across measures to provide a single score. The calculation is the same as for the Threshold Model. The aggregate task score ( $S_t$ ) is given by:

The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is given by:

Pros:

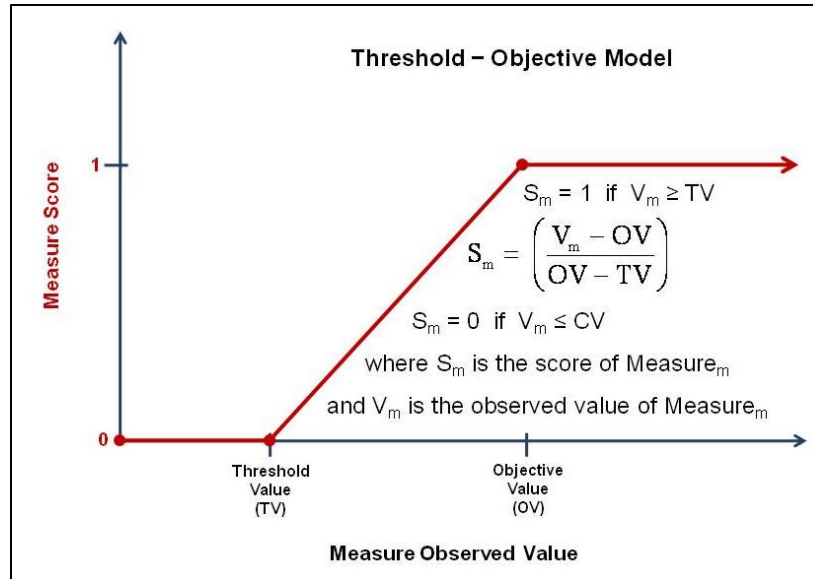
- Good model to show incremental improvements
- Provides a simple linear weighted value
- Provides a continuous linear value added score that includes threshold and objective values
- Scoring not subjective to ordinal scale

Cons:

- An observed value that fails to meet the threshold value may have a minimal impact on the overall score
- A high value above the threshold value for a measure may offset low values below a threshold for other measures
- Improvement across the measured continuum may not be linear

4. **Threshold – Objective Model.** The Threshold – Objective Model incorporates the objective value as a part of the function. This model recognizes value added when a measure's observed value is greater than the threshold value. The additional value is linear based on the threshold and objective values. Any value above the objective value has no incremental improvement. Figure D-3 illustrates the score given to a measure based on the observed value, threshold value, and objective value. Any observed value below the threshold value is scored as a 0. A score of 1 is given to a measure that meets or exceeds an objective value. Any observed value between the threshold and objective values is scored between 0 and 1 based on a linear extrapolation.





**Figure D-3. Threshold - Objective Model**

The model scores can then be aggregated across measures to provide a single score. The calculation is the same as for the Threshold Model. The aggregate task score ( $S_t$ ) is given by:

The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is given by:

Pros:

- Good model to show incremental improvement above the threshold values
- The score is based on a ratio of the observed value to the threshold value
- Provides a single value between 0 and 1
- Places emphasis on meeting and beating threshold values
- Adds emphasis to meeting objective values
- If observed value is significantly higher than threshold value, then less stringent statistical requirements and lower risk

Cons:

- Improvement between threshold and objective may not be linear
- Score of 0 when observed value just meets the threshold value
- Subjective as to value added in score when meet objective value

5. **Threshold – Objective Model 2.** This Threshold – Objective Model 2 uses the same measure scoring function shown in figure D-3. This model also aggregates measures across a single task in the same way as the Threshold – Objective Model. What differs is in the aggregation of those measures for all the tasks in a single mission. This model provides greater emphasis on those measures and tasks that meet or exceed objective values by normalizing across the number of tasks vice the sum of task scores. Since the value of a single task score ( $S_t$ ) can be greater than 1, the overall task performance score for a mission can be greater than 1. The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is given as:

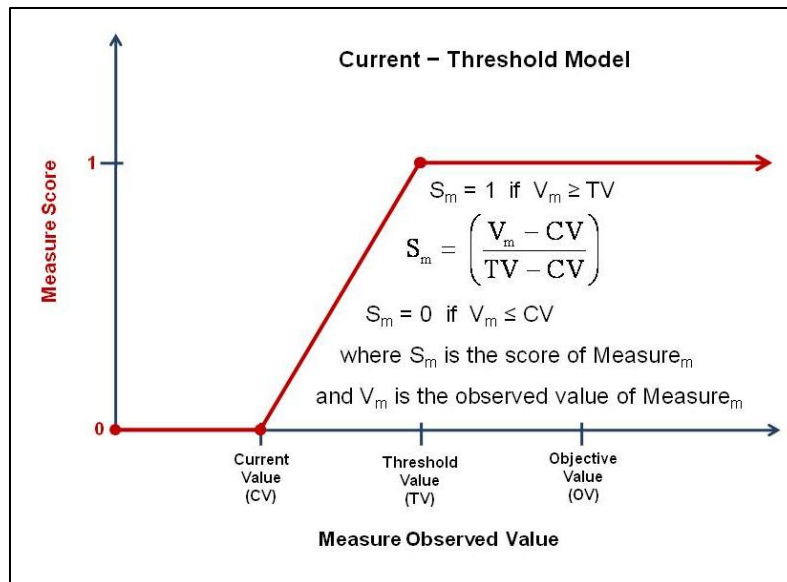
Pros:

- Provides a single value between 0 and 1
- Places increased emphasis on meeting or exceeding threshold values
- Provides equal importance across tasks
- If observed value is significantly higher than threshold value, then less stringent statistical requirements and reduced risk

Cons:

- Improvement between threshold and objective may not be linear
- Subjective as to value added in score when meet objective value
- Score of 0 when observed value just meets the threshold value
- Might skew the scoring upward by assigning greater values to measures and tasks that meet or exceed objective values

6. **Current – Threshold Model.** The Current – Threshold Model incorporates a current value as a part of the function. This model recognizes a current value as a baseline and the threshold value as a level above that as an improvement in capability. The objective value is considered in this model, but does not improve the score above the threshold value. Figure D-4 illustrates the score given to a measure's observed value based on the current value, threshold value, and objective value. Any observed value below the current value is scored as a 0. A score of 1 is given to a measure that meets or exceeds an objective value. Any observed value between the current value and threshold values is scored between 0 and 1 based on a linear extrapolation.



**Figure D-4. Current - Threshold Model**

The CTM scores can then be aggregated across measures to provide a single score. The calculation is the same as for the Threshold Model. The aggregate task score ( $S_t$ ) is given by:

The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is given by:

Pros:

- Good model to show incremental improvement above the current baseline values
- Same as the Threshold Model if the current value is not known and/or is set at the threshold value
- The score is based on a linear interpolated improvement between the current value and the threshold value
- Provides a single value between 0 and 1
- Places emphasis on meeting and beating current values
- Provides an overall score of 1 if all measures meet the threshold values
- Any value above 0 shows improvement from current values

- If observed value is significantly higher than threshold value, then less stringent statistical requirements and lower risk

Cons:

- Improvement between current value and threshold value may not be linear
- May not know current values (Note: Current values can then be set at threshold value.)
- Score of 0 when observed value just meets the current value
- No value added when above the threshold value

7. **Current – Threshold Model 2.** This Current – Threshold Model 2 uses the same measure scoring function shown in figure D-4. This model also aggregates measures across a single task in the same way as the Threshold – Objective Model. What differs is in the aggregation of those measures for all the tasks in a single mission. This model provides greater emphasis on those measures and tasks that meet or exceed objective values by normalizing across the number of tasks vice the sum of task scores. Since the value of a single task score ( $S_t$ ) can be greater than 1, the overall task performance score for a mission can be greater than 1. The overall score of task performance ( $S_p$ ) for all weighted tasks in a mission is given as:

Pros:

- Same as the Current – Threshold Model
- Places equal importance across tasks

Cons:

- Same as the Current – Threshold Model

## ANNEX E

### ASSESSMENT PROCESS MATHEMATICAL MODELS

#### **Mission Effectiveness**

A mission is defined as the task, together with the purpose, that clearly indicates the action to be taken and the reason therefore (Joint Publication [JP] 1-02, *Department of Defense Dictionary of Military and Associated Terms*, November 8, 2010 [As Amended through January 31, 2011]). “The mission establishes the requirement to perform tasks and provides the context for each task’s performance (including the conditions under which a task must be performed). It determines where and when a task must be performed (one or more locations). Finally, it determines the degree to which a task must be performed (implied in the concept of the operation) and provides a way to understand precisely how the performance of a task contributes to mission success (that is, the standard).” (Chairman, Joint Chiefs of Staff Manual [CJCSM] 3500.04E, *Universal Joint Task List [UJTL] Manual*, August 25, 2008).

Based on the definition, a mission has two components to it, task and purpose. A mission is commonly thought of as a set or thread of tasks. The purpose is what needs to be evaluated at the mission (M) level.

The mission purpose can be described by a mission statement, mission objectives, and mission desired effects. “The mission can be described in broad terms by a mission statement. This is a short sentence or paragraph that describes the organization’s essential task (or tasks) and purpose - a clear statement of the action to be taken and the reason for doing so. The mission statement contains the elements of who, what, when, where, and why, but seldom specifies how.” (JP 1-02) “The mission is then further defined with objectives that are clearly defined, decisive, and attainable goals toward which every operation is directed.” (JP 5-0)

A desired effect (DE) can also be thought of as a condition that can support achieving an associated objective, while an undesired effect is a condition that can inhibit progress toward an objective. (JP 5-0, *Joint Operation Planning*, December 26, 2006) An effect is defined as (1) the physical or behavioral state of a system that results from an action, a set of actions, or another effect; (2) the result, outcome, or consequence of an action; and (3) a change to a condition, behavior, or degree of freedom. (JP 1-02) Since capabilities are required to “achieve desired effects,” per the Joint Capabilities Integration and Development System (JCIDS) capability definition, desired effects are what is measured to show mission accomplishment.

#### **Mission Desired Effects**

Each mission has desired effects that are based on mission objectives.

Desired effects may be weighted based on priority in mission objectives and effects.

An assessment of each mission would then be:

Note that mission desired effects are not based on the performers of the mission, but on the recipient (effectee) of the mission. Think in terms that every mission is in support of someone else that benefit or suffer from the mission. Desired effects have attributes that help to define those quantitative and qualitative characteristics that can be measured.

### **Desired Effect Attributes**

A desired effect can be evaluated based on attributes. Attributes will focus on the characteristics of the effect that can be quantified. The attributes should help to “establish the state related to achieving the objectives.” (JP 5-0).

Each attribute should be weighted for a desired effect based on relative importance. For example, is timeliness or accuracy more important in performing the task?

An assessment of a desired effect would then be:

To put in terms of Mission **i**, it can be shown as:

### **Mission Measures**

Desired effects can have measures of effectiveness (MOE) based on the effectees and attributes. MOEs help answer questions like, “Are we doing the right things, are our actions producing the desired effects, or are alternative actions required?” (JP 5-0)

Mission measures may be weighted based on “how well” the measure assesses the desired effect attribute. In many cases, only one measure may exist for a desired effect attribute and, thus, will be weighted a value of 1.0. When two or more measures exist for a single desired effect attribute, there may be a primary measure that will have more weight than the other measures.

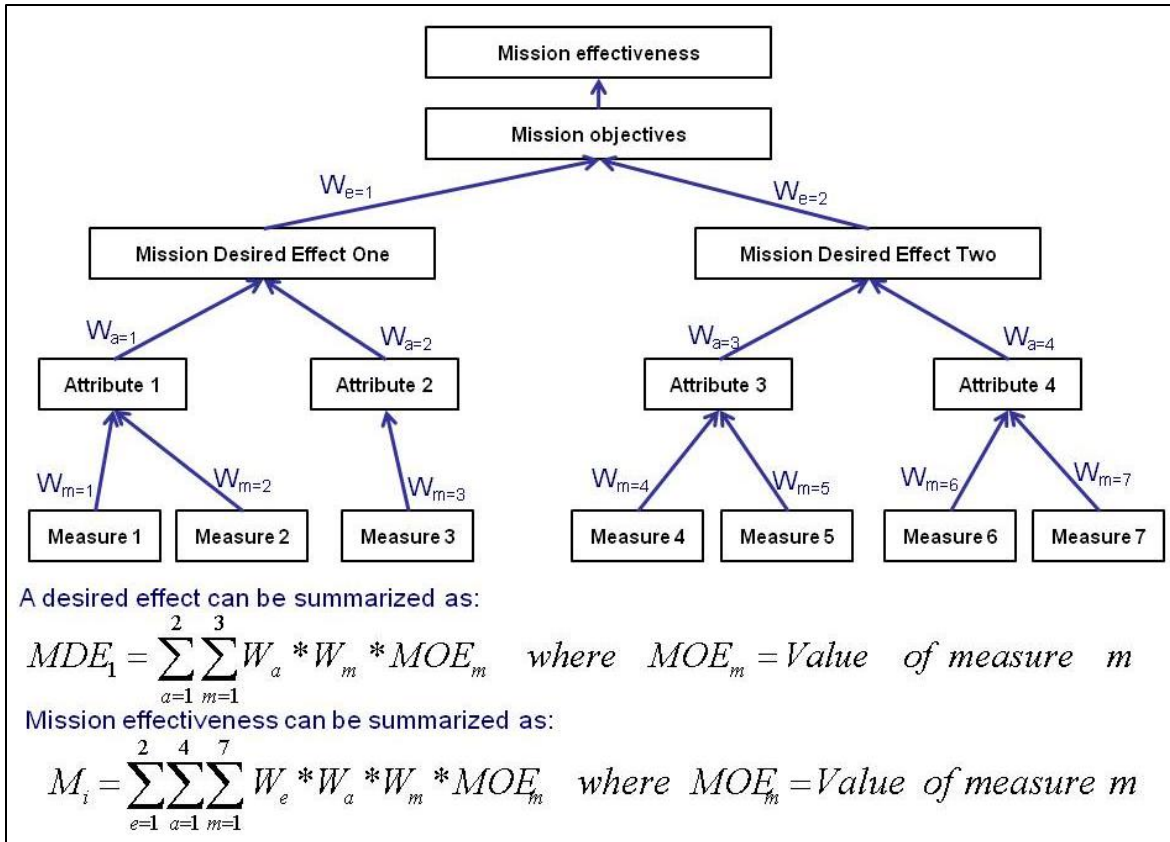
For each Desired Effect Attribute **a**:

Therefore, for each Desired Effect **e**:

and for each Mission **i**:

Figure E-1 provides an example of the evaluation of a mission that is based on mission desired effects, attributes, and measures.

Given that systems may be required to perform several different but related missions, the system under test (SUT) may need to consider each of the missions in the assessment. If a SUT has different missions, then:



**Figure E-1. Example Evaluation of Mission Effectiveness**

To assess the SUT across all its missions may require a weighting of each mission and a breakdown of missions to desired effects and measures. Each mission priority may be based on percentage of system operations in that mission or, based on risk, how often and impact.

Mission effectiveness that is based on all SUT missions **m** may be written as:

In terms of mission measures, overall SUT mission effectiveness can be rewritten as:



## **Critical Operational Issues**

Critical operational issues (COI) are used within the test community to formulate the basis for a test that is focused on operational effectiveness and operational suitability. COIs are the operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined in operational test and evaluation (OT&E) to evaluate and/or assess the system's capability to perform its mission. Effectiveness COIs are typically mission focused and stated as a question, "Can the SUT support the \_\_\_\_\_ mission?"

### **NOTE**

**A critical operational issue (COI) is "a key operational effectiveness (OE) and/or operational suitability (OS) issue (not a parameter, objective, or threshold) that must be examined in OT&E to determine the system's capability to perform its mission. A COI is normally phrased as a question that must be answered in order to properly evaluate OE or OS."**

Since operational effectiveness COIs are based on missions, these COIs may be broken down into sub-COIs that focus on mission desired effects. Although these sub-COIs can be written as questions, it is suggested that sub-COIs take the form of "Assess SUT impact on mission desired effect." Developing sub-COIs that are based on mission desired effects provides options in assessing the mission. In some situations, it may not be feasible to conduct test vignettes that include the system-of-systems (SoS) elements needed to measure mission desired effects. When possible, data should be collected on mission measures that allow a quantitative assessment of the mission. However, the alternative is to make a qualitative assessment of mission desired effects through assessing sub-COIs. Thus, in developing a test report on the SUT, the report will include a written assessment on how the SUT supports each mission desired effect.

### **Example**

The F-35 aircraft is assigned the mission to conduct close air support (CAS). The F-35 will act as the strike aircraft to attack enemy targets for friendly ground forces. The ground force is the beneficiary of the CAS mission with desired effects of: (1) reduce threats, (2) minimize collateral damage, and (3) prevent fratricide. The COI and sub-COIs may be written as:

COI: Can the F-35 support the CAS mission?

Sub-COI: Assess the F-35 impact on reducing threats.

Sub-COI: Assess the F-35 impact on minimizing collateral damage.

Sub-COI: Assess the F-35 impact on preventing fratricide.

## **Conditions Impact on Mission**

The "capability" definition from JCIDS starts out with "the ability to achieve desired effects under a specified set of standards and conditions... ." Standards are those measures and threshold values to which the system or SoS must perform. Conditions are defined as "those variables of an operational environment or situation in which a unit, system, or individual is expected to operate and may affect performance" (JP 1-02).

The *Universal Joint Task Manual* describes conditions further as:

- Conditions should be factors of the immediate environment.
- Conditions should directly affect the performance of a task. A condition must directly affect the ease or difficulty of performing at least one task.
- Conditions should not be a related task.
- Each condition should have a unique, understandable name.
- Conditions may apply to all levels of war and all types of tasks.
- Conditions and descriptors should be written to be compatible with a task-conditions-standards framework. Conditions are expressed within the framework of the phrase, “perform this task under conditions of... .” Therefore, each condition and condition descriptor phrase should fit within this framework.

Conditions must be considered when assessing mission and task performance. Conditions are independent variables of a test. If I apply certain capabilities under a set of conditions, then I can perform my tasks to achieve my desired effects. To consider each condition and its descriptors as a separate test factor would not be feasible. Therefore, conditions must be thought of in terms of sets of conditions that help describe a scenario in which the system or SoS will perform. The Capabilities-Based Assessment process that supports JCIDS indicates that scenarios provide the spectrum of conditions to be considered. Scenarios yield a range of enemies, environments, and access challenges, all of which constitute conditions.

Thus, condition sets need to be identified and established as having certain descriptors to assess system/SoS performance of task and mission in the appropriate operationally realistic environment. These condition sets can be weighted for probability of occurrence. For example, if a system under test is intended to be employed in a desert environment for 75% of its missions, then the condition set related to that scenario should be weighted at 75%. This can be shown as:

To assess the SUT conducting its missions under a set of conditions expected for a given scenario will require a weighting of each mission for that condition set.

Mission effectiveness for single mission may be written as:

Mission effectiveness based on all SUT Missions **m** may be written as:

In terms of mission measures, overall SUT mission effectiveness can be rewritten as:

NOTE: Since the condition set does not affect the weighting of desired effects, attributes, and measures, the sub-script **c** is not included. However, since the measure value may be impacted by the condition **c**, it is included.

### **Task Performance**

The second component of the mission is the set of tasks:

A task is defined as an action or activity (derived from an analysis of the mission and concept of operations) assigned to an individual or organization to provide a capability.

Tasks are performed to enable missions:

Each task should be weighted based on mission critical (or supporting) and system (SUT or not):

Therefore, a mission will be composed of weighted tasks:

Each task has a primary performer and system used in the performance of the task:

Each task has inputs and outputs associated with it. An input/output may be in the form of information exchanges and/or changes in state of an object.

### **Sub-Tasks**

Tasks can be decomposed into sub-tasks, therefore:

### **Task Attributes**

A task can be evaluated based on attributes. Attributes will focus on the output of the task. The attributes should help to answer questions like, “Did the action taken produce results, were the tasks completed to standard, and how much effort was involved?” (JP 5-0). If the task produced an information exchange, then the attributes should help to answer questions like, “ Was the information complete, was the information accurate, and was the information usable?”

Each attribute should be weighted for a task based on relative importance. For example, is timeliness or accuracy more important in performing the task.

An assessment of task performance for a Task **t** would then be:

### **Task Measures**

A task attribute can be same for many tasks. In fact, it is expected sub-tasks will have similar attributes as its parent task. However, its measures may differ for similar task attributes across tasks (sub-task). More prominent will be the fact that measure threshold and objective values will differ for each measure under each task. Thus, for evaluating tasks, measures of performance (MOP) are determined by attributes, but mapped to a specific attribute (task combination).

Each MOP may be weighted in its ability to assess the task attribute. In many cases, only one measure may exist for a task attribute and, thus, will be weighted a value of 1.0. When two or more measures exist for a single task attribute, there may be a primary measure that will have more weight than the other measures.

To assess a single task ( $T_i$ ) based on attributes and measures can be written as:

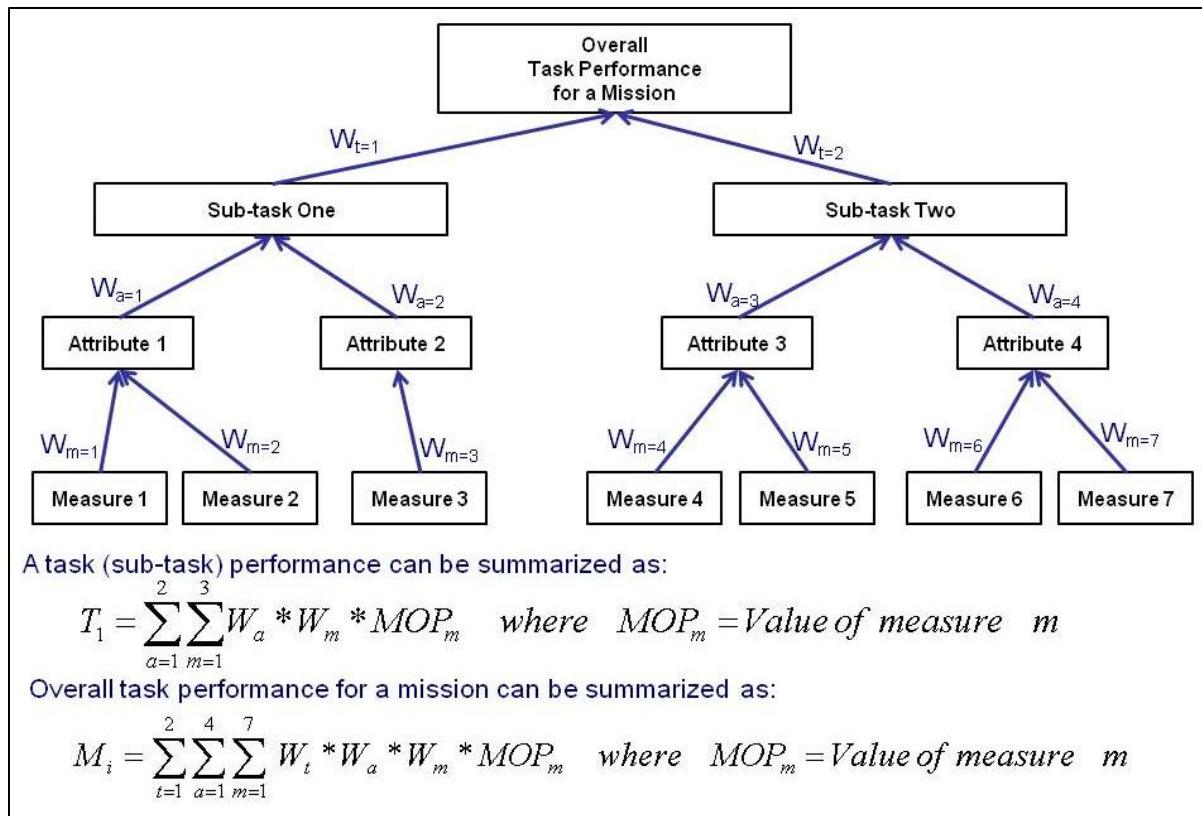
To evaluate task performance across a single mission can be written as:

Figure E-2 provides an example evaluation of overall task performance for a mission that is based on tasks, sub-tasks, attributes, and measures.

#### **Conditions Impact on Tasks**

As stated in the *Universal Joint Task Manual*, “Conditions should be factors of the immediate environment that directly affect the performance of a task.” Similar to the mission, the same condition sets apply to the tasks. These condition sets have the same weight as used for the mission that is based on probability of occurrence. This can be shown as:

To assess the SUT performing tasks under a set of conditions that is expected for a given scenario will require a weighting of each task for that mission and condition set.



**Figure E-2. Example Evaluation of Task Performance**

Task performance for single mission may be written as:

Overall task performance across missions should not be assessed to avoid incorrect conclusions regarding mission effectiveness.